INFECTION NETWORKS, LIFE-HISTORY TRAITS, AND DYNAMICS IN COMPLEX VIRUS-MICROBE SYSTEMS

A Thesis
Presented to
The Academic Faculty

by

Luis Jover

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Physics
Georgia Institute of Technology
August 2016

INFECTION NETWORKS, LIFE-HISTORY TRAITS, AND DYNAMICS IN COMPLEX VIRUS-MICROBE SYSTEMS

Approved by:

Dr. Joshua S. Weitz, Advisor
School of Biology and School of Physics
Georgia Institute of Technology

Dr. Simon Sponberg
School of Physics
Georgia Institute of Technology

Dr. Kurt Weisenfeld
School of Physics
Georgia Institute of Technology

Professor Sam Brown
School of Biology
Georgia Institute of Technology

Dr. Justin Romberg
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: April, 19, 2016

ACKNOWLEDGEMENTS

I would like to thank my advisor, Joshua Weitz, for all his guidance and support throughout my PhD. I feel very lucky to have been part of his group and to have had the chance to interact with so many wonderful people. Special thanks to Michael Cortez, who was also my mentor and collaborator. Mike had to read many of my first drafts, and I can't imagine that being easy.

I would also like to thank current and former group members for enduring my presentations and for their feedback throughout my research : Gabriel Mitchell, Alexander Bucksh, Abhiram Das, Lauren Childs, Charles Wigington, Stephen Beckett, Ceyhun Eksin, Hayriye Gulbudak, Shengyun Peng, Keith Paarporn, Richard In-Ho Joh, Kristen Knipe, Shimantika Sharma.

Special thanks to Joey Leung and Cesar Flores for their collaboration and for the many stimulating conversations. Thanks to Bradford Taylor for putting up with me during the whole PhD experience.

I am also thankful to Justin Romberg for sharing his expertise and advice and to Sam Brown for important insights. I would also like to thank the rest of my thesis committee: Kurt Weisenfeld and Simon Sponberg.

Finally, I would like to thank my mom, my dad and the rest of my family and friends for their unconditional support. Thanks to Kimlee and Eeo for their support and for being part of my adventures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Bacteria and their viral parasites, i.e., phages, are found in natural environments from oceans, soils to the human gut. Phages are key players in ecosystems responsible for a significant portion of microbial mortality. Individual phages can infect a subset of bacteria types in a community as part of complex infection networks. In this thesis we study the interplay between infection networks, life-history traits, and the resulting dynamics in systems with multiple host and phage types. First, we study the trade-off necessary for the coexistence of multiple hosts and phages in systems with statistically realistic infection networks. Second, we study how the trends of network architecture vs. biodiversity depend on life-history traits. Finally, we put forward a method for reconstructing infection networks using measurements of the densities from the dynamics.

CHAPTER I

INTRODUCTION

In 1989 Bergh et al. published a paper that drastically changed the way we perceive the ecological importance of viruses and the role they play at a global scale. The premise of the paper was fairly simple: to count the number of viruses in aquatic environments [5]. The consensus up to that point was that virus particles were not very numerous in the ocean. Presence of viruses in natural water samples was usually attributed to contamination from sewer waters. Surprisingly, Behrg et al. found a concentration of the order of 10^8 virus particles per milliliter. Since then, it has been established that viruses are by far the most abundant biological entities in natural waters and the world. This fact completely changed our view of the ecological potential of viruses and sparked the field of viral ecology [6, 7, 8, 9].

Of course, high number of viruses would mean very little ecologically if they did not interact with other organisms. But viruses, indeed, have a profound effect on most living organism and specifically on microbes. Microbes play a central role in global scale biochemical cycles. In the oceans they are part of the nitrogen and carbon cycle and generate approximately half of the oxygen on Earth [10, 11]. Viruses, in turn, play a key role as agent of microbe mortality and as a consequence are also key players in global geochemical cycles. They are estimated to account for a large portion of bacterioplankton death killing an estimated 10%-40% of the bacteria in the oceans on a daily basis [7, 12]. This has non-trivial consequences for the ecology of microbes in the oceans. Viruses convert bacterial biomass back into dissolved and particulate form, which is not biologically available to higher trophic levels, and is instead available for subsequent growth of the surviving microorganisms [13, 14, 12,

15].

Viruses are also highly diverse. Advances in viral metagenomics have allowed the characterization of community-wide viral genomic composition directly from enviromental samples [16], and show that viral metagenomes are mostly compose of novel viral sequences [17, 18]. This fact remains true even with the steady accumulation of sequenced viral genomes available in searchable databases [19]. In these environmental samples, the majority of viruses are parasites of microbes, whose diversity has been better characterized and is also very high. Thus, in environmental samples you typically find a highly diverse virus-microbe community. This type of systems is the focus of the present thesis. We use theoretical and computational methods to study systems whose characteristics and structure are inspired by naturally occurring systems.

To understand the ecological impact of viruses in microbial communities it is important to characterize which organims they are able to infect. Typically, each virus infects a subset of all the different microbe types in a community. We represent the information of which virus type infect which microbe type with a bipartite infection network. In a bipartite infection network the two types of nodes represent viruses and microbes respectively, and the links represent ability to infect. The pattern of infection networks found experimentally is in general none-trivial. Recent efforts to characterize these patterns have revealed that these complex network have structures that differ statistically from randomly generated networks. Instead, nested infection networks are commonly observed in virus-bacteria systems. In a nested network the host range of a virus and the virus range of hosts form ordered subsets resulting in a hierarchy for who can infect whom [4]. In particular, the specialist virus can infect the most permissive bacteria, whereas the bacteria that is most difficult to infect is infected by the more generalist virus. Nested infection networks in particular and the degree of nestedness of infection networks in general play a important role in

the present thesis. We use nested networks to study systems that resemble naturally ocurringe systems and we use nestedness as a dimension that we can vary to test different hypothesis when varying network structure.

One of the central subjects of this thesis is the relationship between network structure and community dynamics. To study this relationship we rely on theoretical models to describe the interactions between all the different players in a community. Theoretical models of virus-microbes interactions have been used for decades to better understand these systems [20, 21, 22]. One of the earliest models was developed by Levin et al. [21] to study the interaction between bacteria and phage in chemostat experiments. This study focuses on low-dimensional systems with one bacteria strain and one or two types of viruses. In contrast, here we study high-dimensional systems to reflect the type of communities found in environmental samples. To do so, we use a generalized Lotka-Volterra models which include competition between microbe types and infection by viruses. This model has the same mathematical form as the common predator-prey models used in ecology [23, 24]. Other high-dimensional models of virus-bacteria interactions include higher trophic levels [25, 26, 27]. Here, we abstract away other trophic levels to focus on the interactions between viruses and bacteria.

A central theme in this thesis is biodiversity, i.e., the number of surviving species in a community. Throughout this thesis we study what are the conditions promoting biodiversity. This is a central question in ecology and one that has been tackled from many angles and in a variety of different systems [28, 29]. Theoretical models have played an important role in investigating questions related to biodiversity and stability in complex ecological systems. The seminal work by May sparked the debate over complexity vs stability in high-dimensional systems [30]. This debate is still ongoing. The results presented in this thesis contribute to this debate. Although our motivation comes from virus-microbe systems, the results presented here can be easily generalized to other predator-prey systems and help illuminate similar questions that

arise in systems with different types of interactions. For example, nested networks are also found in mutualistic plant-pollinator systems, and questions regarding the role of nestedness and its relationship to life-history traits in promoting coexistence also arise in those systems.

In this thesis we study the interplay between networks structure, life-history traits and community dynamics in virus-microbe systems. Motivated by naturally occurring communities, we start by looking at the specific case of virus-bacteria systems with perfectly nested infection networks and study the conditions that allow the survival of all the species in a community. Then, we extend this question to encompass a ensemble of networks and ask the more general questions of survival vs level of nestedness considering different ranges of life-history traits. Finally, we solve the inverse problem of inferring infection networks from discrete measurements of the dynamics in experiments where mutliple virus and bacteria types interact. This final chapter presents a method with the potential of overcomming limitations of current experimental methods of network inference. These advances would help the field of viral ecology in general and would help validate (or falsify) the conclusions and hypothesis presented in the first two chapters.

In the first chapter we ask if the network property of nestedness might yield insight into mechanisms allowing coexistence in multi-strain host-phage systems. We find general requirements for the coexistence of viral and host strains in perfectly nested systems. First, there should exist a trade-off in the viral strains between life-history traits and host range.These viral traits include: adsorption rate, burst size, and death rate. Second, there should exist a trade-off in the host strains between resistance and growth rate. We find that properties of these trade-offs can predict relative densities of the viral and host strains. We also study systems that are not perfectly nested and discuss how our analysis provide insights in the more general endeavor of understanding the relationship between the structure of networks and

the properties of ecological communities such as biodiversity and stability.

In the second chapter we show how looking at variation in life-history traits is key, although sometimes overlooked, aspect to disentangle the relationship between infection network nestedness and biodiversity. We show how in certain regions of parameter space it is possible to obtain positive relationships between nestedness and biodiversity while at the same time obtaining an opposite trend in different regions of parameter space. We show how this trend are not only local and instead are robust to increases in the focal region of parameter space.

In the third chapter we propose an inverse-based method to infer virus-bacteria infection networks from measurements of host and viral densities in communities where several host and virus types interact. Quantifying who infects whom in a community is essential to understand how the individual-based traits affect ecosystem functions in complex environments. Cross-infection between phage and bacteria has traditionally been determined using culture-based methods, e.g., the plaque assay. Yet the majority of bacteria and phage are currently unculturable. Applying our method we find robust regimes where accurate reconstruction is possible given complex phage-bacteria infection networks with overlapping ranges of infection. We use *in silico* experiments to explore difference characteristics of this method, including the performance given noise in measurements.

CHAPTER II

MECHANISMS OF MULTI-STRAIN COEXISTENCE IN HOST-PHAGE SYSTEMS WITH NESTED INFECTION NETWORKS

Adapted from L. F. Jover, M. H. Cortez, and J. S. Weitz, Mechanisms of multi-strain coex- istence in hostphage systems with nested infection networks, J. Theor. Biol. 332, 65 (2013). [31].

Bacteria and their viruses ("bacteriophages") coexist in natural environments forming complex infection networks. Recent empirical findings suggest that phage-bacteria infection networks often possess a nested structure such that there is a hierarchical relationship among who can infect whom. Here we consider how nested infection networks may affect phage and bacteria dynamics using a multi-type Lotka-Volterra framework with cross-infection. Analysis of similar models have, in the past, assumed simpler interaction structures as a first step towards tractability. We solve the proposed model, finding trade-off conditions on the life-history traits of both bacteria and viruses that allow coexistence in communities with nested infection networks. First, we find that bacterial growth rate should decrease with increasing defense against infection. Second, we find that the efficiency of viral infection should decrease with host range. Next, we establish a relationship between relative densities and the curvature of life history trade-offs. We compare and contrast the current findings to the "Kill-the-Winner" model of multi-species phage-bacteria communities. Finally, we discuss a suite of testable hypotheses stemming from the current model concerning relationships between infection range, life history traits and coexistence

in complex phage-bacteria communities.

2.1 Introduction

Bacteria and their viral parasites, i.e., phages, are found in natural environments from oceans, soils to the human gut. There are an estimated 10^{30} bacteria on Earth, with estimates of phages approximately 10-fold higher [6, 7]. Phages are not only abundant, but they are also key players in ecosystems. For example, phages can be responsible for a significant portion of microbial mortality, e.g., with estimates ranging from 20%-80% [32, 33, 8]. These estimates of lysis are at the community scale. However, individual phages infect a subset of bacteria in a community. A growing number of empirical studies have begun to investigate the nature of cross-infections between phages and bacteria [34, 35, 36, 37]. These studies have the potential to help identify the basis for phage-induced mortality, by delineating the specific phage types capable of infecting and lysing specific host types and, potentially, the taxonomic and biogeographic drivers of cross-infection. Although predictive models of cross-infection remain elusive, it is evident that a single virus can infect multiple strains of a host [38, 3], in some cases multiple species [35, 36], and even hosts from different genera [39].

In an effort to identify cross-system trends and patterns in cross-infection, we (and collaborators) recently re-analyzed 38 phage-bacteria infection studies. This re-analysis identified a recurring pattern in these studies: viral strains have overlapping host ranges such that the overall network of infections is significantly nested [4]. In nested systems, there is a hierarchy for who can infect whom (see Figure 1A). In a perfectly nested system, the specialist virus can infect the most permissive host, the next most specialized virus infects the most permissive host and the second most permissive host, and so on (see Figure 1B). Hence, the host that is most difficult to infect is infected only by the most generalist virus. In a nested system that is

not perfectly nested, this nesting of infection ranges occurs more frequently than expected by chance. It is important to note that some of the nested phage-host infection networks re-analyzed in [4] are derived from experimental evolution studies in which the hosts and phages do not coexist at the same time point, but rather the nested relationship is only observed when performing cross-infection experiments between time points (e.g. [38]). However, some networks are derived from ecological studies where samples are taken from the environment (e.g., [3, 35, 36]). In such cases the finding of overlapping infection ranges poses a dilemma for understanding coexistence. In particular, how can a specialist virus coexist when it only infects a single host, and indeed the host that is most susceptible to infection? Further, how can a permissive host coexist with other more resistant hosts when there is a nested relationship to infectivity? The consequences of such interaction networks on the structure and dynamics of microbes and their viral parasites have not yet been established. Here, we integrate empirical observations of complex infection networks into ecological models of host-phage dynamics.

Host-phage systems and their population dynamics have been studied mathematically for over 30 years [21, 22]. The earliest models were meant to facilitate understanding of relatively low-diversity chemostat experiments, often involving the cross-infection of a single phage type with a single host type. These studies often found that mutants could arise and so simple models were often extended to include two hosts and one phage, two phages and one host, and so on [21, 40]. However, host-phage models applied to natural environments require a greater diversity of bacteria and phage types. There are different types of approaches to integrate diversity into dynamic models. On the one hand, eco-evolutionary models have been developed that assume genotype-phenotype relationships for viral and host traits. Then, mutations can arise *de novo* in the context of an ecological model such that new strains emerge, invade, and possibly lead to diversified communities. Despite their

computational complexity, multiple examples of such eco-evolutionary approaches for phage-host dynamics are now available [41, 42, 43, 44]. Alternatively, ecological models in which a diversity of types are included from the outset can be used to study how certain features in a community may help maintain diversity. Examples of such approaches include spatial multi-strain models [45], models of competition between hosts possessing different types of immune systems [46] and models that incorporate higher trophic levels [25].

The most prominent theory of phage-bacteria ecological dynamics in multi-species communities is of the latter type and is known as the Kill-the-Winner model [25, 26]. In this model, multiple species of bacteria and phage are considered, for which each virus is assumed to exclusively infect a single host type. The model's central conclusion is that each host type is controlled, in a top-down fashion, by a single viral type, with the exception of a single host type whose density is controlled by the total host biomass limit set by an additional generalist grazer. Moreover, the steady state densities of viral types are determined by relative differences in life history traits of the hosts (an issue we return to in the discussion). The model is meant to describe the dynamics of bacterial species, i.e., no strain level dynamics are considered, although its structure is general.

Here we extend the basic framework of the Kill-the-Winner (KTW) model to incorporate complex interaction networks and ask how nestedness mediates coexistence in multi-strain host-phage systems. We focus our attention on the idealized case of a perfectly nested interaction network, and show conditions on the life-history parameters necessary for coexistence. We find trade-off conditions necessary for coexistence and show that coexistence occurs even when the system is perturbed from the equilibrium. We examine the abundances of both hosts and viruses at steady state and their relationship to life history traits and to infection range. Finally we examine the existence of coexistence equilibria in the general case where the infection matrix is

not perfectly nested. We close by discussing the relevance of the current study to the KTW model, empirical efforts to link infection patterns with life history traits, and recent attempts to establish a link between network structure and biodiversity.

2.2 Methods

We model a system of n bacterial strains (hosts) and n viral strains (phages). Bacteria compete for implicit resources while viruses infect different subsets of the bacterial community. We denote the density of host i by H_i and the density of virus j by V_j. Our model of the ecological dynamics of the different host and viral strains is,

$$\begin{aligned} \frac{dH_i}{dt} &= r_i H_i \left(1 - \frac{\sum_{j=1}^{n} H_j}{K} \right) - \sum_{j=1}^{n} M_{ij} \phi_j H_i V_j, \\ \frac{dV_j}{dt} &= \sum_{i=1}^{n} M_{ij} \phi_j \beta_j H_i V_j - m_j V_j. \end{aligned} \tag{2.1}$$

In the absence of viruses, hosts exhibit logistic growth with exponential growth rate r_i and a community-wide carrying capacity K. To simplify the model we assume that for a given host, intrastrain and interstrain competition are the same. The parameters ϕ_j and β_j are the adsorption rate and burst size (virion release per infection) of virus j (which we assume is independent of host i). Virus j decays outside the host at a rate m_j. M denotes the infection matrix, where $M_{ij} = 1$ if virus j can infect host i and $M_{ij} = 0$ if virus j cannot infect host i. For analytical tractability, there are several characteristics of the life cycle of a viral infections that are not included in this model. For example, we assume there is no delay between infection and virion release. We also do not include the possibility of lysogeny, where viral genetic material is incorporated into the host chromosome and vertically transmitted to daughter cells for future activation and lysis.

Empirically measured phage-host infection networks are often statistically nested (for example, see Figure 2.1a). Here, our main goal is to examine the ecological

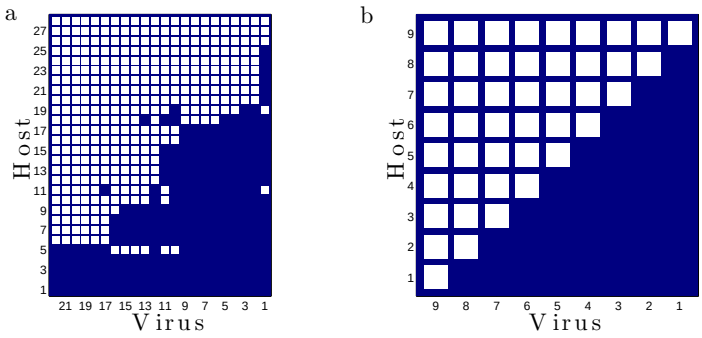

Figure 2.1: (a) Infection network from an experimental study presenting a statistically nested pattern (original data from [3] reanalyzed in [4]). The numbers identify different types of viruses and hosts. (b) Perfectly nested infection network. For a perfectly nested network, the numbers correspond to the rank (i.e number of interactions). White squares denote that a given virus can infect the host.

implications of nested interaction networks for host-phage dynamics. We start by studying the idealized case of a system where the infection network is perfectly nested; see Figure 2.1b. In this representation of the network, viruses are numbered from the most specialist to the most generalist. We will refer to this number as the rank of a species. We see that virus 1 infects only one strain of host whereas virus j infects j different host strains. Note that, in the nested case, virus $j+1$ can infect all strains of hosts that virus j can infect plus an additional one. Similarly, hosts are ordered by the number of viral strains that can infect them. Host i can be infected by i different viral strains. Again, note that in the perfectly nested case, host i can be infected by all viruses that can infect host $i+1$. We relax these simplifying assumptions in section 2.3.4, where we consider the case of interaction networks that are not perfectly nested (as is the case in Figure 2.1a).

Numerical simulations of the ecological dynamics were done using a Runge-Kutta method (ode45 in MATLAB [47]). The parameters (life-history traits) used in all the simulations are shown in table A.1 in A.2. These parameters were chosen from a baseline of biologically realistic values found in the literature [1, 2].

2.3 Results

2.3.1 Equilibrium densities and conditions for coexistence

We start by examining the equilibrium densities of the host and phage strains. When the $2n$ equations of system (2.1) are simultaneously zero, we find the equilibrium densities corresponding to strain coexistence:

$$H_n^* = h_1, \quad H_{n-1}^* = h_2 - h_1, \quad \ldots, \quad H_1^* = h_n - h_{n-1}. \tag{2.2}$$

$$V_n^* = \frac{r_1}{\phi_n}\left(1 - \frac{h_n}{K}\right), \quad V_{n-1}^* = \frac{(r_2 - r_1)}{\phi_{n-1}}\left(1 - \frac{h_n}{K}\right), \quad \ldots \quad V_1^* = \frac{(r_n - r_{n-1})}{\phi_1}\left(1 - \frac{h_n}{K}\right), \tag{2.3}$$

where the superscript * denotes equilibrium densities and $h_j = \frac{m_j}{\phi_j \beta_j}$. The equilibrium densities of the hosts are expressed exclusively in terms of the life-history traits of the viruses (h_j). This is due to the top-down control of the hosts by the viruses. In Figure 2.2 we show graphically how to construct the host steady states. We see that virus 1, which exclusively infects host n, determines the equilibrium density of that host. The value of this density is the ratio $h_1 = \frac{m_1}{\phi_1 \beta_1}$. This result is the same as in a system with just one host and one virus, and it can be interpreted as the host density necessary to support the virus infecting it (Figure 2.2a). Virus 2, on the other hand, requires a host density of $h_2 = \frac{m_2}{\beta_2 \phi_2}$ to persist in the system. This virus infects two hosts: one in common with virus 1 (host n), and an additional one (host $n-1$). However, the density of host H_n is already set by virus 1, so the density of host $n-1$ is the difference between the required density for virus 2 (h_2) and the density of virus 1 (h_1) (Figure 2.2b). Similarly, virus 3 infects one more type of host than virus 2, and thus, the density of the additional host (H_{n-2}) is the difference between the density required by virus 3 to survive (h_3) and the density set by virus 2 (h_2); see Figure 2.2c. The densities of the remaining hosts are determined in an analogous way, Eq (2.2). From this result we obtain that the total host biomass at steady state is $\sum_{j=1}^{n} H_j^* = \frac{m_n}{\beta_n \phi_n}$, i.e. it is set by the most generalist virus.

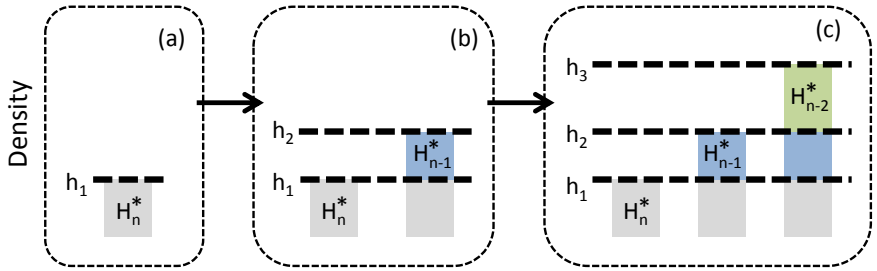

Figure 2.2: Construction of the host steady states (H_i^*). (a) The equilibrium density of host n, H_n^*, is set by the aggregate life-history traits of virus 1, h_1, which infects H_n exclusively. (b) The equilibrium density of host $n-1$, H_{n-1}^*, is set by the difference between h_2 and h_1. (c) The equilibrium density of host $n-2$, H_{n-2}^*, is set by the difference between h_3 and h_2.

Coexistence of all strains at positive densities requires three conditions to be met:

$$\frac{m_n}{\beta_n \phi_n} > \frac{m_{n-1}}{\beta_{n-1}\phi_{n-1}} > \cdots > \frac{m_2}{\beta_2 \phi_2} > \frac{m_1}{\beta_1 \phi_1} \tag{2.4}$$

$$r_n > r_{n-1} > \cdots > r_2 > r_1 \tag{2.5}$$

$$\sum_{j=1}^{n} H_j^* = \frac{m_n}{\beta_n \phi_n} < K. \tag{2.6}$$

These conditions establish a connection between the life-history traits of the different viral and bacterial strains and the structure of the infection networks, specifically the rank of the strains. Equation (2.4) represents a tradeoff for the viruses and can also be written as $h_n > h_{n-1} > \cdots > h_2 > h_1$. Larger values of h_j imply a virus has a higher de-activation rate (higher m_j) and/or produces fewer viruses per infection and/or is worse at attaching to the host (smaller $\beta_j \phi_j$). Thus, the inequalities in equation (2.4) provide the viral trade-offs necessary for coexistence: a virus with a broader host range has less advantageous life-history traits, as characterized by its ratio h_j, compared to viruses with a narrower host range. Note that if two viral types infected a single host, the type with the lower ratio h_j would out-compete the other. Here, coexistence is possible because the viral type with higher h_j can infect more

host strains than the viral types with lower h_j. Equation (2.5) describes the trade-off between immunity and growth rate for the hosts. Specifically, coexistence is possible if a host that can be infected by more viral types has a higher growth rate compared to a host that can be infected by fewer viral types. Finally, equation (2.6) specifies that the sum of the host densities at equilibrium needs to be less than the carrying capacity of the system.

Figure 2.3 shows an example of the dynamics resulting from a system with 2 host and 2 viral strains which satisfies the conditions listed in equations (2.4), (2.5), and (2.6). In contrast, Figure 2.4 shows what happens when the conditions for coexistence are not satisfied. In this example, host 1 has a larger growth rate than host 2, which results in the extinction of host 2, and as a consequence, the extinction of virus 1.

2.3.2 Community dynamics and invasion

Conditions (2.4), (2.5), and (2.6) guarantee the existence of a coexistence equilibrium in system (2.1). In this section we address two issues related to the coexistence between host and viral strains when those conditions are satisfied. First, via numerical simulations we investigate if the species densities tend to the coexistence equilibrium or if cyclic coexistence is also possible. Second, boundary equilibria where one or more of the host and viral strains are extinct also exist in system (2.1). We ask if those boundary equilibria are unstable with respect to invasion by the extinct host and viral strains when conditions (2.4), (2.5), and (2.6) are satisfied.

As seen in Figures 2.3, 2.5, and 2.6, when conditions (2.4), (2.5), and (2.6) are satisfied, the host and viral densities can either tend to steady state or exhibit cyclic oscillations. In Figure 2.3, the coexistence equilibrium point is stable, and the values of the densities tend to the equilibrium values after transient oscillations. For a different set of life-history traits, the coexistence equilibrium point is no longer stable, but cyclic coexistence is still possible (Figure 2.5). Figure 2.6 is an example of cyclic

coexistence for 5 species of bacteria and 5 species of virus that satisfy the trade-off conditions (2.4), (2.5), and (2.6). In Figure 2.6, coexistence is still possible because the coexistence equilibrium exists, even though the time series may be irregular. Note that due to the dimension of the model we do not have a closed form solution for when cyclic dynamics arise in our system.

In appendix A we show that conditions (2.4), (2.5), and (2.6) imply that all boundary equilibrium points of system (2.1) are unstable with respect to invasion by at least one host or viral strain that is absent from that subsystem. This implies that if the dynamics in that subsystem tend to the boundary equilibrium point, then that subsystem can be invaded by one or more of the extinct host or viral strains. A stronger conclusion can be reached if system (2.1) is permanent (i.e. densities are bounded above and, after some time, are bounded below by a finite value [48]). Because the average long term invasability conditions along orbits in permanent Lotka-Volterra systems are equal to the invasability conditions at equilibrium points [48], conditions (2.4), (2.5), and (2.6) imply that the extinct strains can always invade the subsystem when system (2.1) is permanent. It is an open question whether system (2.1) is permanent when conditions (2.4), (2.5), and (2.6) are satisfied.

2.3.3 Relationship between abundance and rank

Another relevant question is the connection between rank and density. Is there a way to infer information about the infection network from measurements of density or vice-versa? We find that, with the exception of host n, host densities are determined by the difference between consecutive h_i values (eq. (2.2)). So, in our framework, measurements of density could inform us about the differences in the life-history traits of viruses with the most similar host range. If two viruses of consecutive rank are very similar (in terms of the aggregate life-history traits h_i), then the corresponding host density is low. On the other hand, two very different values of h_i for consecutive

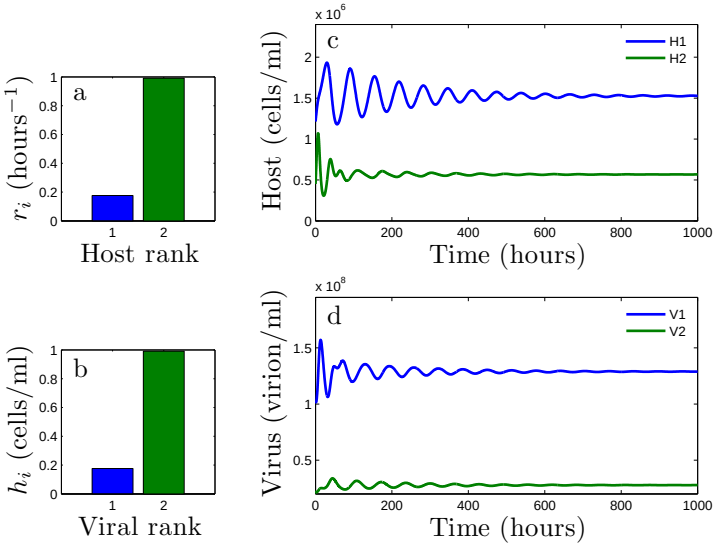

Figure 2.3: Dynamics of 2 host types and 2 viral types when the infection network is perfectly nested and the life-history traits satisfy the conditions (2.4), (2.5), and (2.6) for coexistence. (a) Growth rate, r_i, as a function of host rank in accordance with condition (2.5). (b) h_i as a function of viral rank in accordance with condition (2.4). (c) Host densities as a function of time: both hosts coexist via a stable equilibrium. (d) Viral densities as a function of time: both viruses coexist via a stable equilibrium.

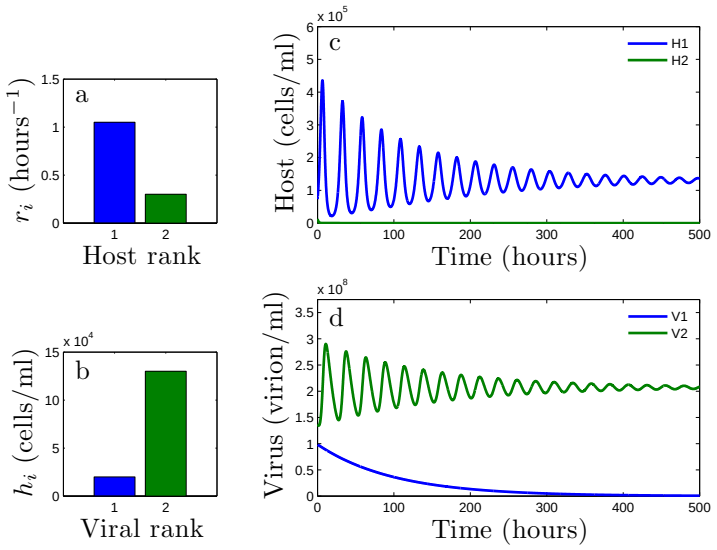

Figure 2.4: Dynamics of 2 host types and 2 viral types when the infection network is perfectly nested and the trade-off conditions are not met. One of the viral types does not survive because the life-history traits do not satisfy the conditions for coexistence. (a) Growth rate, r, as a function of host rank does not satisfy condition (2.5). (b) h as a function of rank. (c) Host density as a function of time: host 2 goes extinct. (d) Viral densities as a function of time: virus 1 goes extinct.

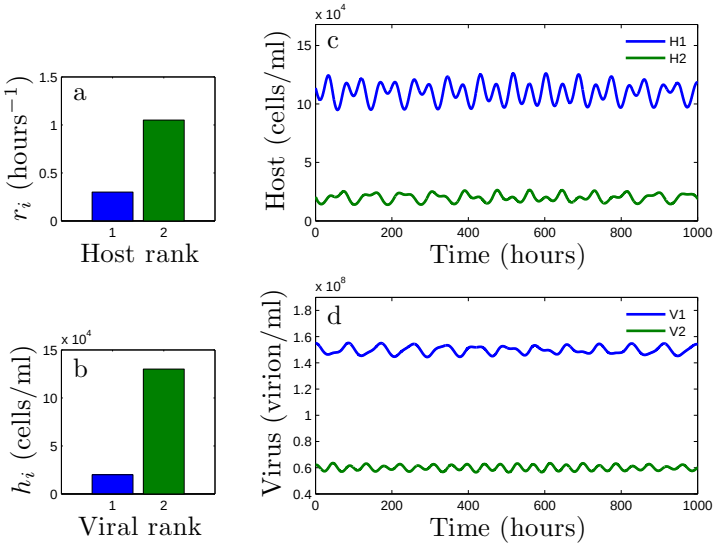

Figure 2.5: Cyclic coexistence of 2 host types and 2 viral types. The infection network is perfectly nested and the life-history traits satisfy the conditions for coexistence. (a) Growth rate, r_i, as a function of host rank in accordance with condition (2.5). (b) h_i as a function of viral rank in accordance with condition (2.4). (c) Host densities as a function of time: both hosts coexist via a stable equilibrium. (d) Viral densities as a function of time: both viruses are present via oscillations.

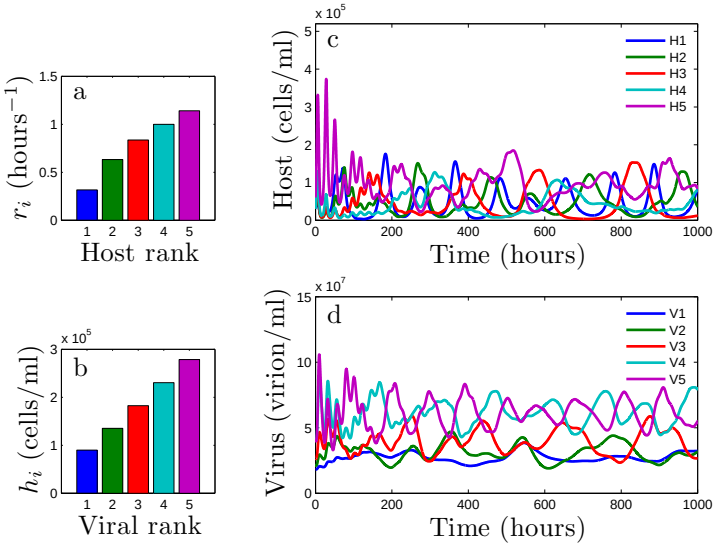

Figure 2.6: Cyclic coexistence of 5 types of hosts and 5 types of viruses. The infection network is perfectly nested and the life-history traits satisfy the conditions for coexistence. (a) Growth rate, r, as a function of host rank in accordance with condition (2.5). (b) h as a function of viral rank in accordance with condition (2.4), (c) host densities as a function of time: all of the hosts coexist, (d) viral densities as a function of time: all of the viruses coexist.

viruses imply high host density. The only exception to this analysis is the density of host n, which is determined uniquely by the ratio h_1 of virus 1 and not by the difference between the ratios of viruses of consecutive rank.

Figure 2.7 shows examples of the connection between the life-history traits of the viruses and the density of the hosts. In the special case where the trade-off curve for the h_i has a curvature with a constant sign, the values of h_i translate into a simple (monotonic) rule for the host densities as a function of rank. For example, when the trade-off between h_i and viral rank is concave up, the H_i^* increase with rank (Figure 2.7a). When the trade-off is concave down, the H_i^* decrease with rank (Figure 2.7b). In both cases, the density H_n^* of host n can be an exception. Figure 2.7c shows the general case. The values of h_i increase with rank following condition 2.4 for coexistence, but the corresponding equilibrium densities need not be a monotonic function of rank.

The viral densities depend, in part, on how different the host strains are. Specifically, the equilibrium densities of the viruses depend on the differences between growth rates of host strains of consecutive rank. However, they don't depend uniquely on the traits of the host, they also depend on the adsorption rate of the focal virus (eq. (2.3)). Thus, information about the viral densities need not translate directly into information concerning differences in the life-history traits of hosts.

2.3.4 Deviation from a perfectly nested network and coexistence

We now return to our assumption about interaction networks and consider systems that are not perfectly nested. The ecological dynamics follow system (2.1), however the different structures of the infection network will be reflected in the matrix M, which includes information of who can infect whom.

The equilibrium densities can be expressed in a compact form as two matrix equations using the infection matrix M and its transpose M^{T}:

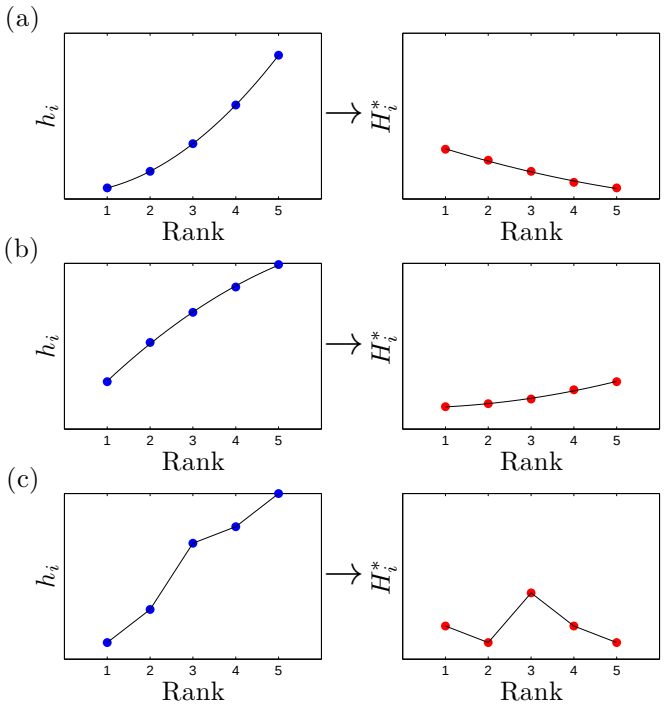

Figure 2.7: Relationship between the viral tradeoff curve (left) and the corresponding host densities at steady state (right). (a) When the trade-off curve is concave up, host density decreases with rank. (b) When the trade-off curve is concave down, host density increases with rank. (c) In general, the host densities are set by the difference of consecutive h_i and need not be a monotonic function of rank.

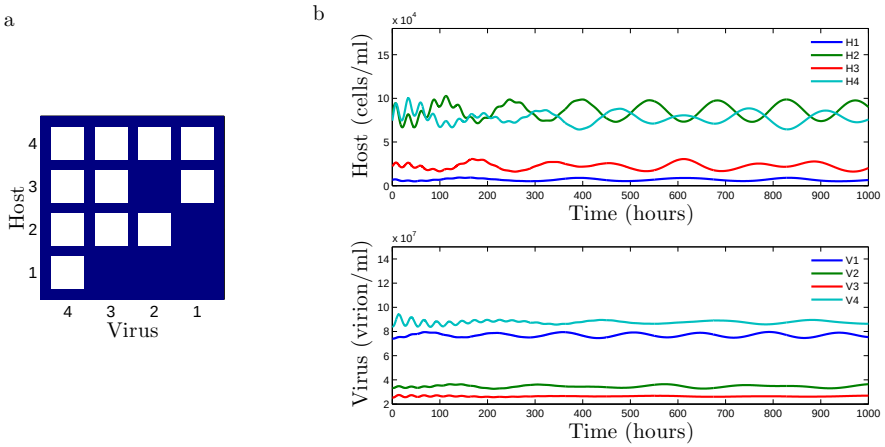

Figure 2.8: (a) Infection networks corresponding to an invertible infection matrix. (b) Time series from numerical simulation using the infection matrix in (a).

$$M^{\mathsf{T}}\vec{H}^* = \vec{h}, \qquad M\vec{V}^{*\prime} = \vec{r}\left(1 - \frac{\sum_{j=1}^{n} H_j^*}{K}\right) \tag{2.7}$$

Here $\vec{H}^*$ and $\vec{V}^*$ are vectors of the equilibrium densities, $\vec{h}$ is a vector whose elements are the h_i, and $\vec{r}$ is a vector whose elements are the r_i. We also use the change of variable $V_i^{*\prime} = \phi_i V_i^*$. We are interested in the solutions of system (2.7) that are positive. We consider two cases: the infection matrix is invertible and the infection matrix is singular. An invertible infection matrix can be interpreted biologically as each viral strain having an unique niche, where the niche is defined by the host range. A singular infection network can be interpreted as the existence of niche overlap between two or more viral strains.

First consider the case where M is invertible, i.e., there is niche differentiation. In this case, there exists a unique solution to system (2.7), and therefore a unique set of coexistence equilibrium densities. The equilibrium densities are expressed in terms of the life-history traits of the different viruses and bacteria and will, in general, involve differences between traits or combinations of traits (e.g. h_i). Therefore, for every invertible infection network there exists a series of inequalities that are necessary

to guarantee positive equilibrium densities (analogous to conditions (2.4), (2.5), and (2.6) for the perfectly nested case). In Figure 2.8a we show an example of an infection network that corresponds to an invertible infection matrix that is not perfectly nested. We also show time series for a system with that specific interaction matrix where all the strains are present in the community (Figure 2.8b). The relationship between network structure and equilibrium densities may be suitable for further numerical analysis.

If the matrix M is not invertible, i.e. there is niche overlap between viral strains, then there are two ways in which coexistence can occur. The first case occurs when the life-history traits of some viral strains are effectively equal. For example, in a system with two viral strains and one host strain, the viral strains can coexist if they have the same h_i values. Biologically, this case is unlikely to occur given that we are describing strains in terms of function, and in this case, coexistence is possible only when strains are functionally identical.

The second case in which coexistence is possible occurs if we relax assumptions about the adsorption rate and the burst size for viral strains. In our formulation, a specific viral type infects all host strains with the same adsorption rate and burst size. However, if viruses exploit host strains at different rates, i.e., adsorption rate and burst size depend on both viral and host type, then coexistence is possible even with complete niche overlap. In this case, the steady states are solutions of two analogous systems of equations where the infection matrix M is a weighted matrix whose entries correspond to the adsorption rate and burst size for each host-viral interaction. A limited number of studies have shown that viral infection rates can differ significantly between host strains (e.g., see [4]). Hence, biologically, the use of quantitative information for host-phage infection assays represents an important target for future analysis and may shed light on the drivers of coexistence in natural communities.

2.4 Discussion

We studied the ecological dynamics of phages and hosts using a Lotka-Volterra framework that incorporated complex cross-infection networks. We found that coexistence is possible even when viruses exploit overlapping ranges of host. In the case of a perfectly nested infection network, Figure 2.1b, we found trade-offs for both the hosts and the viral strains that are necessary to allow the coexistence of all the species. The trade-off for hosts implies that host growth rate decreases with defense (the larger the growth rate, the larger the viral range). The trade-off for viruses implies that viruses must be less efficient at utilizing host resources as they increase their host range. We also showed that the densities of the host strains at equilibrium are determined by how different the viruses are in terms of their aggregate life-history traits.

The idealized case of a perfectly nested network is mathematically tractable and may help to identify potential principles underlying coexistence in host-phage systems. On the other hand, real infection networks are rarely perfectly nested (see Figure 1 for an example, and the re-analysis of Flores et al [4]). We showed that coexistence is possible in the current framework if there exists a partitioning of niches. In those cases, the resulting tradeoffs between infection range and life history traits are not easily presented in a general way because each infection network results in different trade-off conditions. This, in turn, makes it difficult to study the stability for the general case. Nevertheless, we showed numerical examples of coexistence away from equilibrium for a infection matrix that is not perfectly nested.

The current model is similar in spirit to the Kill-the-Winner (KTW) model [25]. The KTW model also proposed mechanisms by which bacteria could coexist, stabilized by the presence of viruses. However, in the KTW model, infections are one-to-one, meaning that each bacterial strain can only be infected by one virus and likewise, each virus can only infect one bacteria. This would be equivalent to an diagonal infection network in our representation. In the KTW model, coexistence of different types

of bacteria is achieved through what is known as a "killing the winner" mechanism, where "coexistence among bacteria is ensured by host-specific viruses that prevent the best bacterial competitors from building up" [25]. Hence, viruses enable coexistence in a system that would, in their absence, lead to a diversity collapse. The principle of top-down control also applies to the model presented here, despite the fact that we considered complex interaction networks with the possibility of overlapping host range. Another similarity between the KTW model and the current model is that the viral steady states densities are set partially by the difference in growth rate of different hosts, and are inversely proportional to the adsorption rates of the viruses.

However, the current model makes predictions not found in the KTW model. First, KTW assumes each bacterial strain is infected by a single viral type and so an ordering of bacterial growth rates is not necessary for coexistence. In the current model, an ordering of bacterial growth rates is in fact required for coexistence and represents a prediction of the model. Second, whereas KTW assumes no difference in viral life history traits, we again predict that differences in life history traits are required for coexistence when viruses differ in their host range. Hence, altogether we predict that there should be an entanglement between network structure and bacterial/viral life history traits. We suggest that experimental studies of cross-infection should move beyond qualitative analysis (i.e., whether or not a phage infects a bacteria) to quantitative analysis (i.e., the lysis rate of bacteria by phages). Doing so would help identify costs of resistance and costs of infectivity in natural populations. There can be costs to hosts and to viruses in natural populations that differ in their infection ranges. These costs can include changes in life history traits [49], but also constraints on the degree of resistance/infectivity that a strain can achieve [50]. However, it is important to recognize that there may be other factors affecting resistance/infectivity and infection ranges such as co-evolution of host and viral strains and the spatial structure of the community.

Altogether, the current study predicts that phage-host coexistence in a given system depends on both network structure (e.g., nestedness) and life history traits. Other types of infection networks are possible including modular [4] and multi-scale [51], suggesting the need for further investigations on the relationship between cross-infection and life history traits. However, as we discussed earlier, the present model does not explicitly incorporate (co)evolution. Indeed, our group, and others, have studied coevolution of phage-host theoretically [41, 42, 43, 44, 52]. Experimental studies of coevolving phage-host systems suggest that infection networks are not static, but rather are dynamic and reflect the changing identities of strains in a population [53, 54, 55]. In extending the current model, we point out that coevolution occurs over spatially extended domains. Spatial structure is thought to stabilize diverse interactions amongst phage and bacteria (e.g. [45, 1, 56]). Moreover, geographic structure can play a key role in affecting the outcome of coevolution [57]. Further work is warranted to quantify how structure in infection networks are driven by and act as drivers of the spatial distributions of diverse communities of phages and bacteria [58, 59, 51].

CHAPTER III

MULTIPLE REGIMES OF ROBUST PATTERNS BETWEEN NETWORK STRUCTURE AND BIODIVERSITY

Adapted from L. F. Jover, C. O. Flores, M. H. Cortez, and J. S. Weitz, Multiple regimes of robust patterns between network structure and biodiversity, Scientific reports 5 (2015).

Ecological networks such as plant-pollinator and host-parasite networks have structured interactions that define who interacts with whom. The structure of interactions also shapes ecological and evolutionary dynamics. Yet, there is significant ongoing debate as to whether certain structures, e.g., nestedness, contribute positively, negatively or not at all to biodiversity. We contend that examining variation in life history traits is key to disentangling the potential relationship between network structure and biodiversity. Here, we do so by analyzing a dynamic model of virus-bacteria interactions across a spectrum of network structures. Consistent with prior studies, we find plausible parameter domains exhibiting strong, positive relationships between nestedness and biodiversity. Yet, the same model can exhibit negative relationships between nestedness and biodiversity when examined in a distinct, plausible region of parameter space. We discuss steps towards identifying when network structure could, on its own, drive the resilience, sustainability, and even conservation of ecological communities.

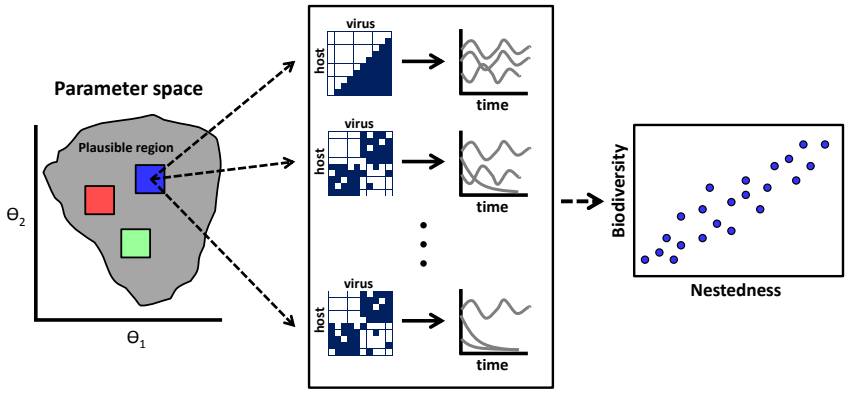

Figure 3.1: Schematic of how studies using numerical simulations have drawn insights into the relationship between network structure and biodiversity. (Left panel) A focal region of parameter space is examined (the blue square, in the parameter space of $\theta_1 - \theta_2$). (Middle panel) Given parameter variation in a focal region, the network is modified across a spectrum of configurations, here from low to high nestedness from bottom to top (where white cells denote interactions and blue cells denote the absence of interactions). Then, the dynamics of each system are simulated and/or analyzed given variation in network structure and model parameters. The proportion of surviving species in the simulations determine the biodiversity of each community. (Right panel) Variation in resulting biodiversity is compared to the variation in network structure, e.g., nestedness (right panel). Given the large number of parameters, such studies do not exclude the possibility that distinct structure-biodiversity relationships may exist for different life history parameters (the red and green boxes in the left panel).

3.1 Introduction

Ecological communities are often composed of a large number of interacting types, e.g., species, morphs or strains. Interaction patterns in a community can be represented as a network where nodes denote distinct types and links between nodes denote connections between individuals of the respective types. These connections can represent distinct modes of ecological interactions including predation, mutualism, competition and parasitism. Understanding the relationship between network structure and the subsequent dynamics of populations has, for decades, been facilitated by theory. For example, Robert May's seminal work in the early 1970s introduced the idea that large complex networks were more likely to be unstable [30, 60]. Whether or

not complexity begets instability in an ecological community remains controversial, as May's original conclusions depend, in part, on assumptions regarding the choice of the underlying interactions and *random* network structure [61, 62, 63, 64, 65].

In reality, ecological interactions are both complex and structured. Realized networks may differ in terms of their connectance, nestedness and modularity - as but three examples of differentiating features. Connectance is defined as the ratio of realized links to potential links; nestedness quantifies the extent to which there exists a hierarchy such that interaction ranges of increasingly specialized types are organized as proper subsets of the interaction ranges of more generalist types [66]; and modularity quantifies the extent to which organisms tend to interact within densely connected groups rather than between groups [67, 68]. Network features, including connectance, modularity, and nestedness [29, 69, 4] have been shown, in theory, to affect the biodiversity and stability of the underlying community [28, 29].

For example, Bascompte and colleagues showed that mutualistic plant-pollinator networks tend to be nested and conjectured that nestedness increases biodiversity, quantified in terms of the relative fraction of surviving types in a dynamic model [28]. Yet, in contrast, James and colleagues [70] argued that nestedness is a covariate of, rather than a driving factor for, increases in biodiversity. Similarly, Suweis et. al showed that elevated nestedness can emerge as a consequence of adaptations that increase species-level and total community abundance [71]. As a second example, phage-bacteria infection networks are often significantly nested [4]. Phage-bacteria communities with nested networks have been shown to be stable [31], so long as there are trade-offs between interaction ranges and other life history traits. Nestedness may even facilitate the emergence of increased biodiversity of both bacteria and phage given a single limiting resource for bacterial growth when contrasted to the dynamics of bacteria-only environments [72].

How is theory used to study the relationship between network architecture and

biodiversity? By and large, theoretical studies typically represent an ecological community via a system of differential equations that describe the change in the population abundances of interacting types. In some instances, it is possible to determine the relationship between network architecture and biodiversity [72]. Yet, when analytical solutions are not available, then many studies select model parameters from biologically plausible, prior distributions and simulate system dynamics to identify statistical relationships between network structure and biodiversity (see Figure 3.1 and [29, 73, 70]). Such approaches raise a question: how does the selection of the prior parameter distribution influence model dynamics and the resulting relationship between network structure and biodiversity? Further, is it possible that there are distinct, robust relationships between network structure and biodiversity given different parameter assumptions?

In this manuscript, we examine the entanglement of network architecture and model parametrization and their combined effect on biodiversity (Figure 3.1). To do so, we simulate ensembles of model dynamics given distinct assumptions of plausible life history traits. We find that the relationship between patterns of network architecture and ecological dynamics can vary qualitatively with model parameterization. Throughout, we focus on a specific model of virus-bacteria dynamics and study the relationship between one ecological property – community biodiversity, i.e., the fraction of surviving strains – and one network property – nestedness. Nonetheless, we explain how our findings can be applied to other systems in which there is uncertainty with respect to the life history traits of interacting strains.

3.2 Results

3.2.1 A rule-based framework to identify distinct domains of biodiversity-nestedness relationships

We are interested in studying coexistence of multiple strains in virus-bacteria systems and its relationship to model parametrization. We define biodiversity as the fraction of

surviving strains in the system and simulate the interactions between strains using the model described in the Methods. In general, coexistence of all the strains in a system, i.e., a biodiversity value of 1, requires the existence of a steady state with positive densities for all strains. For a given interaction matrix, we refer to a steady state with positive densities for all strains as a feasible steady state and to the associated parameter set as a feasible parameter set.

The mathematical conditions for feasibility can be formulated in terms of the parameters of a model: r, bacterial growth rate; K, carrying capacity; a, bacterial competition; as well as ϕ, m, and β; denoting virus adsorption rate, decay rate, and burst size respectively. As an example, consider a low-dimensional system with two bacteria, two viruses, and a nested infection network (virus 1 infects bacteria 1 and 2 and virus 2 infects only bacteria 1). To simplify the analysis we assume that virus adsorption rates and burst sizes are independent of bacterial strains (i.e $\phi_{ij} = \phi_j$, $\beta_{ij} = \beta_j$) and that intra-specific and inter-specific competition between bacterial strains is equal (i.e. $a_{ii'} = 1$). Under these assumptions the conditions for feasibility are: $r_1 > r_2$ and $K > \frac{m_1}{\beta_1\phi_1} > \frac{m_2}{\beta_2\phi_2}$. These conditions represent a feasible "volume" in the 9-dimensional parameter space, i.e., a region of parameter space where coexistence of all 4 strains is possible. Further, we see that the condition for feasibility involving growth rates divides the corresponding two-dimensional parameter subspace into two areas: a feasible region where coexistence of all strains is possible($r_1 > r_2$), and an infeasible region where coexistence of all strains is not possible ($r_2 < r_1$). As should be evident, randomly sampling parameter space would affect conclusions regarding the total potential biodiversity in the system.

We can extend the analysis of this low-dimensional example to the case of a perfectly nested infection network with 10 bacteria and 10 viruses (Figure 3.2a). The rules for feasibility are an extension of the rules from the previous 2 virus, 2 bacteria example, namely: $r_1 > r_2 > ... > r_{10}$ and $K > \frac{m_1}{\beta_1\phi_1} > \frac{m_2}{\beta_2\phi_2} > ... > \frac{m_{10}}{\beta_{10}\phi_{10}}$. These

rules can be generalized for a perfectly nested network of arbitrary size with equal number of bacterial and virus strains [31]. In the case of 10 bacteria and 10 viruses, the rules define a feasible region in the 41-dimensional parameter space of the model. We sample parameters from two different regions of parameter space to illustrate the effect of model parameterization on biodiversity-nestedness relationships. First, we sample parameters from the feasible region of the perfectly-nested network (Figure 3.2a) where all strains coexist when the interaction matrix is nested. Second, we sample parameters from the feasible region of a low-nestedness network (Figure 3.2b); details of the network and its feasible region are presented in the Methods for the rule-based framework.

Given the sampling regions, numerical simulations of model dynamics are averaged over an ensemble of 100 infection networks representing a gradient in nestedness while conserving the connectance of each network in the ensemble. When parameters are chosen from the feasible region of the perfectly nested network, we observe a positive biodiversity-nestedness relationship (Figure 3.3a). In contrast, when parameters are chosen from the feasible region of a low-nestedness network, we see a negative biodiversity-nestedness relationship (Figure 3.3b). The perfectly nested network was included in the ensemble of 100 networks. Sampling parameters from its feasible region resulted in survival of all the strains in the system (average biodiversity of 1, Figure 3.3a). The low-nestedness network used to obtain the second sampling region was not included in the ensemble because it has a difference connectance than the perfectly nested network and the rest of the networks. We found that the trends in biodiversity were robust to changes in initial conditions when sampling values from biologically plausible regions. The key point of this analysis is that the relationship between biodiversity and nestedness differs qualitatively when examining two distinct feasible sets of parameters.

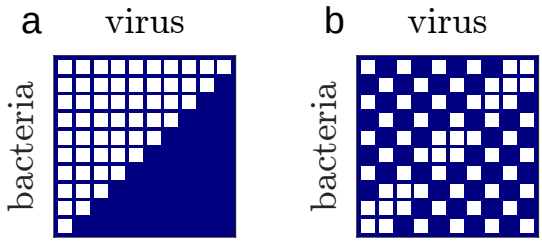

Figure 3.2: Focal matrices used in the rule-based framework. a) Low nested network (NODF = 0.21). b) Perfectly nested network (NODF = 1). White cells denote that a virus in that column is able to infect the bacteria in that row, whereas blue cells denote the absence of infection.

3.2.2 A feasibility-based framework to identify distinct domains of biodiversity vs. nestedness

For a general interaction network the conditions for feasibility are more complicated than the conditions presented for the perfectly-nested network derived in the previous section. As a consequence, we developed an alternative framework for selecting parameter sets that maximize biodiversity in Eqs. (3.1)-(3.2) for any non-trivial infection network. This alternative framework does not require finding the rules for feasibility explicitly. Instead, we choose a subset of the parameters randomly from a biologically plausible region (Table 3.3), specify the target steady state densities of the bacteria and viruses, and then solve for the rest of the parameters using the steady state equations. In this way it is possible to obtain a particular feasible parameter set for any infection network for which all nodes have at least one link (see Supplementary Information for details).

We selected three particular parameter sets for which biodiversity is maximized for three different infection networks of 10 bacteria and 10 viruses. We calculated average biodiversity for an ensemble of 100 matrices which included the three focal networks (Figure 3.4). Figure 3.4 shows that not only is it possible to maximize biodiversity for different networks, but it is also possible that the resulting trends are qualitatively different. We show that, by maximizing biodiversity for a low-nestedness

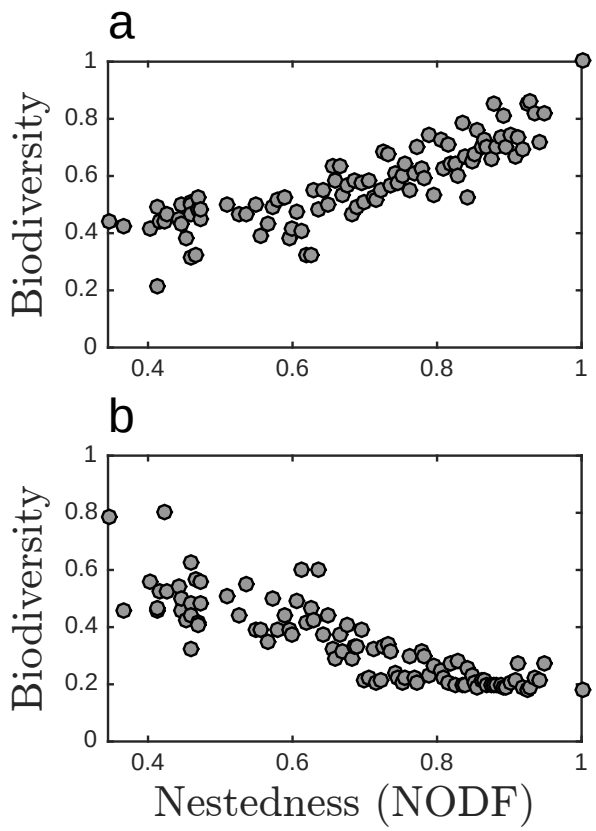

Figure 3.3: Average biodiversity over 100 different sets of parameters for 100 different matrices spanning nestedness values from 0.35 to 1. (a) Parameter sets sampled from the feasible region of the perfectly nested networks (Table 3.1). (b) Parameter sets sampled from the feasible region of a low-nestedness network (Table 3.2).

matrix, we obtain a negative trend of biodiversity vs. nestedness (Figure 3.4 left). In contrast we obtain a positive trend of biodiversity vs. nestedness by selecting parameters that maximize biodiversity for a perfectly nested networks (Figure 3.4 right). It is also possible that the quantitative strength of the relationship can differ, approaching the case where the is no significant relationship between biodiversity and nestedness (Figure 3.4 middle). This analysis supports the prior conclusions, i.e., that biodiversity-nestedness relationships depend on model parameterization.

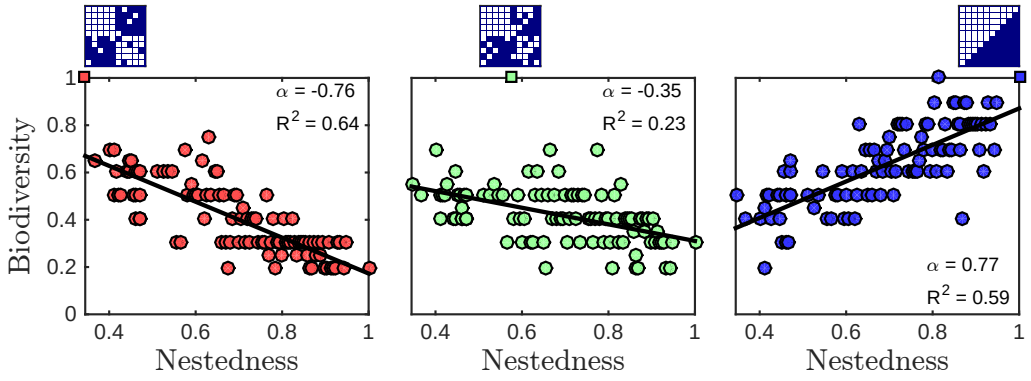

Figure 3.4: For fixed parameter values in different regions of parameter space, the relationship between biodiversity and nestedness can be positive, negative, or exhibit a weak trend. Each panel corresponds to a different (fixed) parameterization of the model. Each point represents the biodiversity for a particular interaction matrix whose nestedness (NODF) lies between 0.35 and 1. The matrix for which biodiversity is maximized is plotted above each panel. (Left, red) A negative biodiversity-nestedness relationship arises when the parameter set maximizes biodiversity for a network with low nestedness (NODF = 0.35). (Middle, green) A weaker negative trend arises when the parameter set maximizes biodiversity for network with intermediate nestedness (NODF = 0.57). (Right, blue) A positive biodiversity-nestedness relationship arises when the parameter set maximizes biodiversity for a network with high nestedness (NODF = 1). The slope of a linear fit α and coefficient of determination R^2 are presented ($p < 10^{-5}$ for all the fitted lines).

3.2.3 Biodiversity-nestedness relationships are robust to local perturbation of parameter sets

Here, we examine biodiversity-nestedness relationships when parameters and network structure are both varied. To do so, we sampled parameter values from regions of parameter space centered about the original parameter set which maximized biodiversity for a particular interaction network (as explained in the previous section). The size of the region was set by the parameter δ such that for a given value of a parameter, $\hat{\theta}$, the random parameter values were sampled from the interval $[\hat{\theta}(1-\delta), \hat{\theta}(1+\delta)]$. Here θ is a dummy parameter denoting any of the life history traits, r, K, m, β and ϕ that impact Eqs.(3.1)-(3.2). Figures 3.5a-f show evidence for a robust, negative biodiversity-nestedness relationship that persists even as the interval width increases.

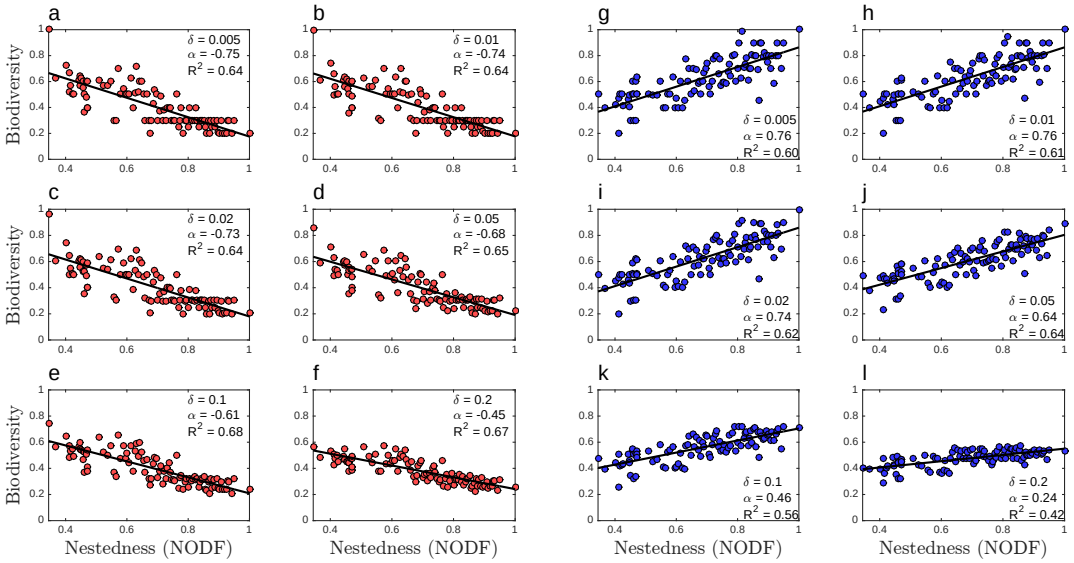

Figure 3.5: Biodiversity-nestedness relationships are robust to local perturbations of parameter values. (Panels a-f, red) Average biodiversity for 100 parameter sets sampled from a uniform distribution centered around a parameter set that maximizes biodiversity for a low-nestedness network. (Panels g-l, blue) Average biodiversity for 100 parameter sets sampled from a uniform distribution centered around a parameter set that maximizes biodiversity for a perfectly nested network. Each plot corresponds to a different value of δ which determines the size of the region used for sampling. The slope of a linear fit α and coefficient of determination R^2 are presented ($p < 10^{-5}$ for all the fitted lines).

We see that the positive trend is robust to selecting parameters from a region, but the trend becomes weaker as the size of the intervals are increased. Similarly, Figures 3.5g-l show evidence for a robust, positive biodiversity-nestedness relationship observed in a different area of parameter space also persists as the sampling interval is increased. In summary, qualitatively distinct biodiversity-nestedness relationships persist, i.e., are robust, when parameter values are chosen at random from a region of parameter space.

3.3 Discussion

We analyzed a nonlinear model of phage-bacteria dynamics (Eqs. (3.1)-(3.2)) as a means to investigate the entanglement between interaction network structure, life history traits, and biodiversity. Given the complexity in the space of possible networks, we considered ensembles of networks that varied in a particular structural feature – nestedness – such that interaction ranges differ in the extent to which they form partially ordered subsets of one another (see Figure 3.7). We found that there is not a globally applicable, positive relationship between biodiversity and nestedness in this model (Figure 3.3). Instead, we identified distinct, robust regions in parameter space where there are contrasting relationships, both positive and negative (Figures 3.4-3.5).

Elevated nestedness is a common feature of interaction networks spanning both plant-pollinator and phage-bacteria systems [69, 4]. Moreover, prior theoretical work has suggested that ecological communities whose interaction networks are nested are more likely to have higher relative biodiversity [69]. Our results highlight the need to understand the life history context underlying a given biodiversity-nestedness relationship. The totality of parameter space includes a subspace of biologically plausible values. Such subspaces often have relatively uninformative prior distributions. Therefore, using biologically plausible regions to restrict the parameters of a model is not a strong enough restriction to uniquely define the effect of network structure on community dynamics.

Recently, it has been pointed out that different model parameterizations can lead to different biodiversity levels and consequently to contradictory results [74]. We expand on this point to show that this is also the case for whole regions of parameter space. This is important because studies using numerical simulations often average over different parameterizations. Indeed, Rohr et al. [74] examine to what extent parameters can vary for a given network structure before the community make-up

changes. In contrast, our work examines how changes in network structure affects community make-up for a fixed set (or region) of parameters. These approaches are related, but they are not the same. Our results show how averaging over parameterizations is not sufficient to account for the effects of life-history traits and that a more systematic study of parameter space is necessary. Additionally, we make a stronger case for the effect of parametrization by showing that it is possible to not only maximize biodiversity for specific networks but to obtain completely different trends of biodiversity vs. nestedness (Figure 3.6).

In our view, statistical inference from numerical simulations can be informative and even advantageous, so long as certain precautions are kept in mind. The key point is that systematic analysis of parameter space is necessary whenever a numerical approach is used to characterize network structure-biodiversity relationships in nonlinear ecological systems. Studies that rely on analytical methods to estimate biodiversity or related features usually focus on fixed-point equilibrium states that are stable either locally or globally. This could be problematic for two reasons. First, general analytical solutions could be hard to find or interpret. Second, coexistence in high-dimensional ecological models could be characterized by non-equilibrium steady states. In such circumstances, fixed-point analyses would overlook configurations that are ecologically relevant.

In the case of phage-bacteria dynamics, our study highlights the value of additional measurements of life history traits, complementary to the recent focus on methods for characterizing who infects whom [75]. We used a particular phage-bacteria model to highlight the importance of systematically studying model parametrization in distinct regions to better understand the relationship between network structure and biodiversity - yet the findings are relevant to a wider debate. The current findings point to the need to revisit the relationship between network structure, life history traits and biodiversity in other systems and given other kinds of network patterns.

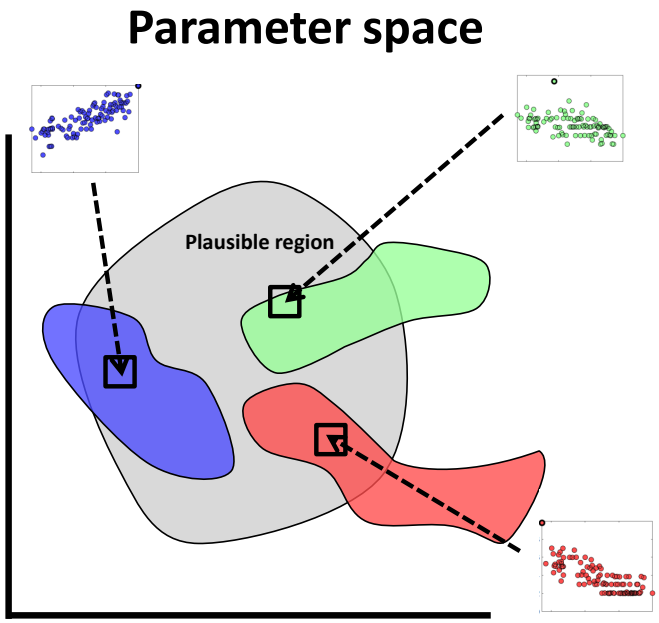

Figure 3.6: Schematic representation of parameter space: biologically plausible region and feasible regions for different infection networks. Black squares represent sampling regions that result in different trends of biodiversity vs. nestedness.

Optimistically, the systematic study of model parametrization could be of service in resolving ongoing debates concerning the relationship, or relationships, between biodiversity and network structure in plant-pollinator systems [28, 70, 76, 74].

3.4 Methods

3.4.1 Model

The dynamics of virus-bacteria systems can be modeled using systems of coupled, nonlinear differential equations [21, 22]. Here we use a system of equations that extend the basic Lotka-Volterra equations [77, 24] to incorporate multiple types of

Table 3.1: Parameter ranges used to obtain feasible parameter sets for the perfectly nested network. Parameters were sampled from uniform distributions and sorted according to the rules of feasibility for the perfectly nested network presented in rule-based framework of the Results. The limits of the distributions are specified in the column titled Range. A fixed value was used for K.

Parameter (units)	Range\Value
r_i (1/d)	3.61 - 43.35
ϕ_j (ml/(virus $\cdot$ d))	2.4×10^{-7} - 2.4×10^{-6}
β_j (virus/cell)	10 - 100
m_j (1/d)	0.037 - 0.52
K (ml)	$max\left(\frac{m_i}{\beta_i \phi_i}\right) \times 10 = 2.17 \times 10^6$

Table 3.2: Parameter ranges used to obtain feasible parameter sets for the low-nestedness network. Range denotes the limits of the uniform distribution used to generate parameters. The limits were calculated such that the parameters satisfy the rules of feasibility for the low-nestedness network presented in the feasibility-based framework as shown in the Methods. Fixed values were used for ϕ, β, and K.

Parameter (units)	Range\Value
r_i for $i = 1, 2, 3, 5, 6, 8, 9, 10$ (1/d)	4.06 - 4.51
r_i for $i = 4, 7$ (1/d)	3.61 - 4.06
ϕ_j (ml/(virus $\cdot$ d))	1.32×10^{-6}
β_j (virus/cell)	30
m_j for $j = 1, 2, 3, 5, 6, 8, 9, 10$ (1/d)	0.47 - 0.52
m_j for $j = 4, 7$ (1/d)	0.42 - 0.47
K (ml)	$\frac{m_4+m_7}{\beta\phi} \times 10 = 2.36 \times 10^5$

virus and bacteria and competition between bacteria [31]:

$$\frac{dH_i}{dt} = r_i H_i \left(1 - \frac{\sum_{i'}^{n_h} a_{ii'} H_{i'}}{K}\right) - H_i \sum_{j}^{n_v} M_{ij} \phi_{ij} V_j, \tag{3.1}$$

$$\frac{dV_j}{dt} = V_j \sum_{i}^{n_h} \beta_{ij} \phi_{ij} M_{ij} H_i - m_j V_j, \tag{3.2}$$

where there are n_h bacteria types, each with density H_i and n_v virus types, each with density V_j. In this system: r_i is the growth rate of bacteria i in the absence of phage and other bacteria, $a_{ii'}$ is the competitive effect of bacteria i' on bacteria i (set to 1 for the analysis), K is the system-wide carrying capacity, ϕ_{ij} is the adsorption rate of

Table 3.3: Parameter and target steady state density ranges used in the feasibility-based framework. Bacteria growth rates, r_i, and virus decay rates, m_j, were derived using the steady state equations and the parameters presented in this table (see Methods, given feasibility-based framework). The range denotes the limits of the uniform distributions used to generate parameters.

Parameter (unit)	Range\Value
ϕ_j (ml/(virus· d))	2.4×10^{-8} - 2.4×10^{-7}
β_j (viruses/cell)	10 - 100
H_i^* (cell/ml)	10^3 - 10^4
V_j^* (virus/ml)	10^6 - 10^7
K (ml)	$max(H_i^*) \times 10 = 10^5$

phage j when attaching to bacteria i, β_{ij} is the effective burst size of phage j when infecting bacteria i, m_j is the decay rate of virus j. Finally the element M_{ij} denotes which virus infects which bacteria such that $M_{ij} = 1$ if bacteria i is infected by virus j and is zero otherwise; altogether these interactions can be represented as a network (Figure 3.7) where each type is equivalent to a strain [78].

We are interested in identifying fixed points where all strains have positive densities, such that:

$$(\beta \circ \phi \circ M)^T \vec{H} = \vec{m} \tag{3.3}$$

$$(\phi \circ M)\vec{V} = \vec{r}\left(1 - \frac{\sum H_{i'}}{K}\right) \tag{3.4}$$

We refer to these points as feasible fixed points. In the present model the feasible fixed point is the solution $\vec{H}^*$ and $\vec{V}^*$ to Eqs. (3.3) and (3.4). Here, $\vec{r}$ and $\vec{m}$ are vectors whose elements are the growth and decay rates and β, ϕ, and M are matrices whose elements are the β_{ij}, ϕ_{ij}, and M_{ij} respectively; "$\circ$" denotes element-wise matrix multiplication.

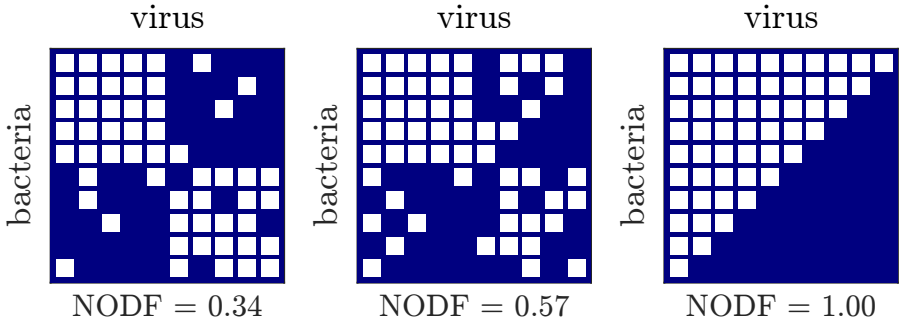

Figure 3.7: Focal matrices used in the feasibility-based framework. The left-most network has low nestedness and the rightmost network has high nestedness. Each row represents the interactions of a given bacteria type with all viruses and each column represents the interaction of a given virus type with all bacteria. White cells denote instances in which a given virus in column j infects a given bacteria in row i (i.e., $M_{ij} = 1$), whereas blue cells denote the lack of infection (i.e., $M_{ij} = 0$).

3.4.2 Network ensemble

We generated ensembles of networks with fixed connectance and fixed size, $n_h \times n_v = n^2$, given the further assumption that $n = n_v = n_h$. The ensemble generation procedure builds upon that introduced by James and colleages [70]. Starting with a perfectly nested network, randomly chosen interactions are removed from the existing interactions among bacteria and viruses and an equal number of new interactions are added among bacteria-virus pairs where interactions did not exist originally. In the conventional matrix representation this is akin to random removal of interactions in the top left corner of the matrix followed by an equivalent addition of new interactions in the bottom right corner of the matrix. Figure 3.7 shows an example of three matrices with different values of nestedness generated using this procedure. To generate a large ensemble of matrices with varying degrees of nestedness the number of randomly selected links that where moved was varied from 1 to $(n/2)^2$. We selected 100 invertible matrices for our ensemble and used the Non-overlapping and Decreasing Fill (NODF) metric for nestedness [66].

3.4.3 Estimating biodiversity from numerical simulation of the dynamics

Biodiversity is defined in this study as the fraction of surviving strains in Eqs. (3.1)-(3.2) after a sufficiently long transient. We estimated biodiversity by numerically integrating the system using MATLAB's ODE45 and enumerating the number of strains with densities greater than 10^{-10} ml at the end of the simulation. We used this criteria for survival instead of relying on stability analysis because strain coexistence can occur via a stable steady state, periodic cycles, or chaotic dynamics.

The stopping time of the simulation was determined via a heuristic that evaluates the convergence of time-averaged densities and mitigates inconsistencies introduced by arbitrarily choosing a stopping time. In developing the heuristic, we leveraged the fact that the average density of each strains is equal to its equilibrium density in Lotka-Volterra systems. The algorithm to determine stopping time is as follows:

- Every 500 hours calculate the infection matrix of only those strains with densities greater than 10^{-10}.
- If the subsystem is solvable (invertible community matrix) calculate the theoretical interior fixed point.
- Check if the average density over the last half of the simulation is within 10 percent of the theoretical prediction of the subsystem.
- The simulation stops when the last condition is satisfied or the simulation has run for 40000 hours.

Note that the algorithm only calculates the stopping time and does not alter the densities of the strains. To evaluate the robustness of the stopping criteria, we calculated persistence using the stopping time heuristics and compared the results with ones obtained after doubling the time for all simulations. Figure 3.8 shows the relationship between persistence and nestedness and the difference in mean persistence.

Note that the relationship between biodiversity and nestedness are the same whether using the heuristic stopping time or twice this value (Figures 3.8a and 3.8b). We find that there are less than 1% differences in the average biodiversity between the two stopping times (Figure 3.8c). Altogether, the heuristic captures the long term coexistence of different strains even when the dynamics do not tend to a steady state.

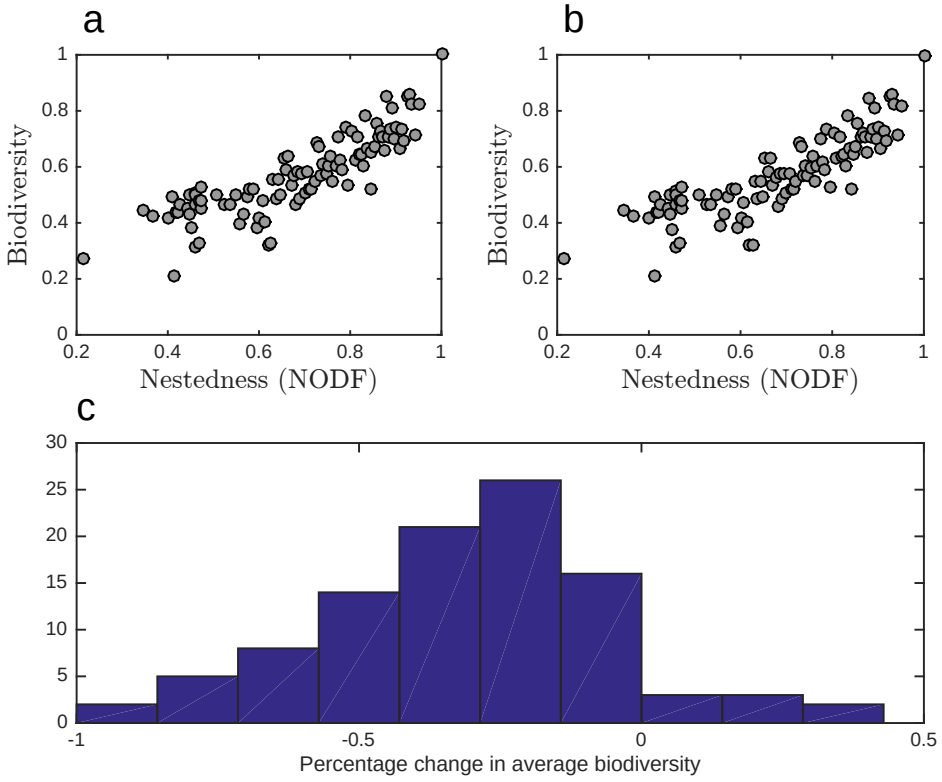

Figure 3.8: Doubling the time of the simulations obtained with the stopping time heuristic results in less than a 1% change in the computed average biodiversity. (a) Average biodiversity using the stopping time heuristic. (b) Average biodiversity using double the time used in (a). Percentage change between (a) and (b).

3.4.4 Parameter range selection

We use two different frameworks to select parameter ranges. In doing so, we use the term "plausible" to denote those parameter sets and steady-state densities whose values are consistent with virus-host biology and ecology. In addition, we use the term

"feasible" to denote those parameter sets whose associated steady state densities are all positive for a given infection matrix. The steps to generate parameter sets in each of the frameworks are:

Ruled-based framework:

- Choose a focal matrix.
- Solve for the feasible steady state of the focal matrix, implicitly in terms of model parameters.
- Choose plausible regions in parameter space that satisfy the feasible steady state conditions.
- Generate a parameter set by sampling uniformly from feasible and biologically plausible regions.

Feasibility-based framework:

- Choose a focal matrix.
- Sample one set of values each of carrying capacity, adsorption rate, burst size, virus and host steady state densities (K, ϕ_j, β_j, H_i^*, V_j^*) for all species from biologically plausible regions.
- Solve the steady state equations to find specific values of host growth rate and virus mortality rate, r_i and m_i, respectively for all species.
- If the values of r_i and m_j are plausible, then these together with the previously sampled K, ϕ_j, and β_j form a feasible and plausible parameter set.
- Generate a parameter set by sampling uniformly from a region centered around the specific feasible parameter set.

We used three focal networks in the feasibility-based framework (Figure 3.7) and two focal networks in the ruled-based framework (Figure 3.2). The following sections explain each framework in detail.

3.4.4.1 Parameter range selection in the ruled-based framework

In the rule-based framework, we start with a particular focal matrix and solve analytically for the steady states. From the steady states we obtain constraints on the life-history traits (rules in the form of inequalities) which guarantee feasibility. For example, feasibility in the perfectly nested network requires an ordering of host growth rates ($r_1 > r_2 > ... > r_{10}$) amongst other rules. We then choose biologically plausible regions of parameter space that satisfy these rules and look at how biodiversity varies with network nestedness. For this, we draw uniformly distributed parameters from these fixed regions and calculate average persistence for an ensemble of networks that span a large range of nestedness values. We followed this framework for two different focal matrices: one with high nestedness (Figure 3.2b) and one with low nestedness (Figure 3.2a). This framework has the advantage that the rules for feasibility characterize the entirety of parameter space by dividing it into feasible and non-feasible regions for each focal network. Nonetheless, it is not trivial to generalize this framework to an arbitrary focal network. As a consequence we were not able to find a low-nestedness network with the same connectance as the rest of the matrix in the ensemble and with simple feasibility rules. Instead, the focal low-nestedness network used to show a different trend of persistence vs. nestedness has a slightly larger connectance (three more interactions) than the rest of the matrices in the ensemble.

3.4.4.2 Feasibility conditions for the perfectly nested network

We make the following assumptions regarding Eqs.(1)–(2): there is equal intra-specific and inter-specific competition across bacterial strains ($a_{ii'} = 1$), and virus adsorption rates and burst sizes are independent of bacterial strains ($\phi_{ij} = \phi_j$, $\beta_{ij} = \beta_j$). Using

these assumptions we find general expressions for the steady states, H^* and V^*, in terms of the life-history traits for a 10 by 10 perfectly nested network (Figure 3.2b) by solving Eqs. (3.3) and (3.4):

$$\vec{H}^* = \begin{pmatrix} \frac{m_{10}}{\beta_{10}\phi_{10}} \\ \frac{m_9}{\beta_9\phi_9} - \frac{m_{10}}{\beta_{10}\phi_{10}} \\ \frac{m_8}{\beta_8\phi_8} - \frac{m_9}{\beta_9\phi_9} \\ \frac{m_7}{\beta_7\phi_7} - \frac{m_8}{\beta_8\phi_8} \\ \frac{m_6}{\beta_6\phi_6} - \frac{m_7}{\beta_7\phi_7} \\ \frac{m_5}{\beta_5\phi_5} - \frac{m_6}{\beta_6\phi_6} \\ \frac{m_4}{\beta_4\phi_4} - \frac{m_5}{\beta_5\phi_5} \\ \frac{m_3}{\beta_3\phi_3} - \frac{m_4}{\beta_4\phi_4} \\ \frac{m_2}{\beta_2\phi_2} - \frac{m_3}{\beta_3\phi_3} \\ \frac{m_1}{\beta_1\phi_1} - \frac{m_2}{\beta_2\phi_2} \end{pmatrix} \qquad \vec{V}^* = \begin{pmatrix} \frac{r_{10}}{\phi_1} \\ \frac{r_9 - r_{10}}{\phi_2} \\ \frac{r_8 - r_9}{\phi_3} \\ \frac{r_7 - r_8}{\phi_4} \\ \frac{r_6 - r_7}{\phi_5} \\ \frac{r_5 - r_6}{\phi_6} \\ \frac{r_4 - r_5}{\phi_7} \\ \frac{r_3 - r_4}{\phi_8} \\ \frac{r_2 - r_3}{\phi_9} \\ \frac{r_1 - r_2}{\phi_{10}} \end{pmatrix} \left(1 - \frac{\frac{m_1}{\phi_1\beta_1}}{K}\right) \tag{3.5}$$

In order for all strains to have positive densities, the life-history traits must satisfy:

$$K > \frac{m_1}{\beta_1\phi_1} > \frac{m_2}{\beta_2\phi_2} > \ldots > \frac{m_9}{\beta_9\phi_9} > \frac{m_{10}}{\beta_{10}\phi_{10}} \tag{3.6}$$

$$r_1 > r_2 > \ldots > r_9 > r_{10} \tag{3.7}$$

These expressions generalize to perfectly nested networks of any size [31].

3.4.4.3 Feasibility conditions for the low-nestedness network

Here we derive constraints on the life-history traits for a low-nestedness network. In order to obtain rules that are easier to interpret we set $a_{ii'} = 1$, $\beta_{ij} = \beta$, and $\phi_{ij} = \phi$. These assumptions imply that: intra-specific and inter-specific competition between bacterial strains is equal and the virus burst size and adsorption rates are independent of host and virus strain. We use the low-nestedness matrix presented in Figure 3.2a,

which has several symmetries that result in simpler rules. By solving Eqs. (3.3) and (3.4) with this matrix we find the following steady states:

$$\vec{H}^* = \begin{pmatrix} m_{10} - m_4 \\ m_9 - m_7 \\ m_8 - m_4 \\ 4m_7 + m_4 - (m_1 + m_3 + m_5 + m_9) \\ m_6 - m_4 \\ m_5 - m_7 \\ 4m_4 + m_7 - (m_2 + m_6 + m_8 + m_{10}) \\ m_3 - m_7 \\ m_2 - m_4 \\ m_1 - m_7 \end{pmatrix} \left(\frac{1}{\beta\phi}\right), \tag{3.8}$$

$$\vec{V}^* = \begin{pmatrix} r_{10} - r_4 \\ r_9 - r_7 \\ r_8 - r_4 \\ 4r_7 + r_4 - (r_1 + r_3 + r_5 + r_9) \\ r_6 - r_4 \\ r_5 - r_7 \\ 4r_4 + r_7 - (r_2 + r_6 + r_8 + r_{10}) \\ r_3 - r_7 \\ r_2 - r_4 \\ r_1 - r_7 \end{pmatrix} \left(1 - \frac{m_4 + m_7}{K}\right) \left(\frac{1}{\phi}\right) \tag{3.9}$$

In order for all strains to have positive densities, the life-history traits must satisfy:

$$
\begin{aligned}
&m_{10}, m_8, m_6, m_2 > m_4 \\
&m_9, m_5, m_3, m_1 > m_7 \\
&4m_7 + m_4 > m_1 + m_3 + m_5 + m_7 \\
&4m_4 + m_7 > m_2 + m_6 + m_8 + m_{10} \\
&r_{10}, r_8, r_6, r_2 > r_4 \\
&r_9, r_5, r_3, r_1 > r_7 \\
&4r_7 + r_4 > r_1 + r_3 + r_5 + r_9 \\
&4r_4 + r_7 > r_2 + r_6 + r_8 + r_{10} \\
&m_4 + m_7 < K
\end{aligned}
\tag{3.10}
$$

3.4.4.4 Parameter range selection in the feasibility-based framework

In this framework parameters are selected from intervals centered around particular feasible parameter sets of a focal network. We make the following assumptions to the general model (Eqs. (1)–(2)): equal intra-specific and inter-specific competition across bacterial strains ($a_{ii'} = 1$), and virus adsorption rates and burst sizes are independent of bacterial strains ($\phi_{ij} = \phi_j$, $\beta_{ij} = \beta_j$). These assumptions are not necessary to implement this framework, but we use them because they simplify the analysis. In particular, under these assumptions, a feasible coexistence equilibrium is guaranteed to exist whenever the infection matrix M is invertible. When these assumptions are not satisfied, one must check whether the matrices $\beta \circ \phi \circ M$ and $\phi \circ M$ are invertible; defined in Eqs. (3.3) and (3.4). If the system is not solvable then the life-history traits generated contain degeneracies that make some combination of viral or host strains effectively equal. We selected three focal infection networks M (Figure 3.7), and values K, ϕ_j, β_j, H^*, V^* from random uniform distributions in biologically plausible regions. Then, using Eqs. (3.3) and (3.4), we obtained values

for m_i and r_i. With this procedure we obtain a set of life-history traits values that yield a feasible fixed point. For each parameter value, θ, local perturbations are made by sampling from an interval of length $2\delta\theta$ centered around θ. The particular values of delta used are included in the caption of Figure 5.

CHAPTER IV

INFERRING PHAGE-BACTERIA INFECTION NETWORKS FROM TIME SERIES DATA

In communities with bacterial viruses (phage) and bacteria, the phage-bacteria infection network establishes which virus type infects which host types. The structure of the infection network is a key element in understanding community dynamics. Yet, this infection network is often difficult to ascertain. The plaque assay is the gold-standard for establishing who infects whom in a community. This culture-based approach does not scale to environmental samples with increased levels of phage and bacterial diversity, much of which is currently unculturable. Here, we propose an alternative method of inferring phage-bacteria infection networks. This method uses time series data of fluctuating population densities to estimate the complete interaction network without having to test each phage-bacteria pair individually. We use *in silico* experiments to analyze the factors affecting the quality of network reconstruction and find robust regimes where accurate reconstructions are possible. In addition, we present a multi-experiment approach where time series from different experiments are combined to improve estimates of the infection network and mitigate against the possibility of evolutionary changes to infection during the time-course of measurement.

4.1 Introduction

Bacterial viruses are ubiquitous and play an important ecological role at the global scale. In the oceans, viruses are responsible for a significant fraction of bacterial mortality and as a result have an effect on global geobiochemical cycles [14, 12,

7, 13]. By killing bacteria, they redirect resources from higher trophic levels and back into the microbial resource pool. Yet, not all bacteria types are susceptible to all virus types. Each phage type potentially infects subset of hosts which can be presented as complex networks of infection [78]. Quantifying who infects whom remains essential to understanding how individual-based traits affect ecosystem-wide properties in complex environments.

For more than 50 years, the host range of phage, i.e., the types of host that a phage type infects, has been measured using plaque assays [79]. A plaque assay is an experimental method in which a growing culture of bacteria on an agar surface are exposed to phage. Clear "plaques" are formed whenever the phage can infect and lyse the target host. Plaque assays are considered the gold-standard for determining infection but are hard to scale-up to community levels. The principal reason is that the majority of phage and bacteria in a community sample are not yet available in culture. In response, a number of (partially) culture-independent methods have been proposed, including viral tagging [80, 81], PhageFISH [82], polonies [9]. Each of these methods requires some degree of culturing or co-visualization of labeled particles, which also presents challenges for scaling-up to complex communities.

Here, we use time series measurements of host and virus densities in experiments where several host and virus types interact to infer the entire community infection network.

The inference of interaction networks from system dynamics is a field of study with wide-spread applications, from inference of gene regulatory network [83, 84], and chemical reaction [85], to neural networks [86]. The key insights from one class of inference methods is that statistical patterns in dynamics, including cross-correlation and mutual-information, can be leveraged to infer interaction [87]. However, such correlation-based approaches can be of limited value when applied to high dimensional systems with nonlinear interaction. As an alternative, Shandilya et al. [88]

showed a method for reconstructing interaction networks from discrete measurements of the time series in systems where the underlying functional form of the interactions is known. Similarly, Stein et al. [89] following the work of Monier et al. [90] used discretized Lotka-Volterra equations to estimate interaction networks, model parameters, and time dependent perturbations in competitive microbial communities. In this study we extend this approach to phage-bacteria systems with antagonistic predator-prey interaction. We lay the theoretical foundations of the method and test its validity using *in silico* experiments. As we show, inferring phage-bacteria infection networks may be possible given appropriate deployment of existing technologies already available to estimate genotype densities over time.

4.2 Method

4.2.1 Model

We model the interaction between N_h host types and N_v virus types using a generalization of the Lotka-Volterra predator-prey equations [77, 24]. The densities of multiple host and virus types are described by a system of differential equations that include the effect of competition between host types and the infection of host by multiple virus types [31, 9]:

$$\frac{dh_i}{dt} = r_i h_i \left(1 - \frac{\sum_{i'}^{N_h} a_{ii'} h_{i'}}{K} \right) - h_i \sum_{j}^{N_v} M_{ij} \phi_{ij} v_j, \tag{4.1}$$

$$\frac{dv_j}{dt} = v_j \sum_{i}^{N_h} \beta_{ij} \phi_{ij} M_{ij} h_i - m_j v_j, \tag{4.2}$$

The model consists of N_H equations of the form (4.1) for the density of each host type, h_i, and N_V equations of the form (4.2) for the virus densities, v_j. In this system: r_i is the growth rate of host i in the absence of viruses and other hosts, $a_{ii'}$ is the competitive effect of host i' on host i, K is the system-wide carrying capacity, ϕ_{ij} is

the adsorption rate of virus j when attaching to host i, β_{ij} is the burst size of virus j when infecting host i, m_j is the decay rate of virus j. Finally M_{ij} is the infection matrix, i.e., a matrix representation of the infection network, which takes a value of 1 if host i is infected by virus j and zero otherwise.

4.2.2 Numerical simulations of the dynamics; infection network ensembles and model parameters

To study the performance of our reconstruction method, we simulated time series of systems where several hosts and virus types interact. We used MATLAB's ODE45 to numerically integrate systems of equations of the form described in Section 4.2.1. In doing so, we utilize both random infection networks and nested infection networks. Nested interaction networks are commonly observed in culture-based analyses, such that the host range of phage and the phage range of hosts form ordered subsets [4]. Following Jover et al. 2015 [91] we generated an ensemble of 100 infection matrices, each one with 10 host types and 10 virus types, spanning a spectrum of nestedness values. The infection matrices were generated by starting with a modular matrix and shifting interactions, through a random process, to regions that increase nestedness [91]. We also found feasible parameter sets (i.e., parameters with positive steady state densities) for each one of the infection matrices. We followed the procedure described in [91] to find feasible parameter sets. Namely, we select a subset of the model parameters and target densities (Table 4.1) and use the steady state equations to solve for the rest of the parameters obtaining a feasible parameter set.

4.2.3 Infection network reconstruction

Our method for reconstructing infection networks requires discrete measurements of the dynamics resulting from the interaction of different host and virus types. This method extends the approach described in [89] to host-phage systems. We will use only the equations describing the dynamics of the viruses (equations of the form

Table 4.1: Parameter and target steady state density ranges used to find feasible parameter sets. Bacteria growth rates, r_i, and virus decay rates, m_j, were derived using the steady state equations and the parameters presented in this table (see Methods, given feasibility-based framework). The range denotes the limits of the uniform distributions used to generate parameters.

Parameter (unit)	Range\Value
ϕ_j (ml/(virus· d)	10^{-8} - 10^{-7}
β_j (viruses/cell)	10 - 50
H_i^* (cell/ml)	10^3 - 10^4
V_j^* (virus/ml)	10^6 - 10^7
K (ml)	$max(H_i^*) \times 100 = 10^6$

(4.2)). . We start by rewriting equation (4.2) in the form:

$$\frac{d\ln(v_j)}{dt} = \sum_i^{N_h} \beta_{ij}\phi_{ij}M_{ij}h_i - m_j. \tag{4.3}$$

We assume that we have $N+1$ measurements of the densities of all virus and host types in the system at times $[t_1, t_2, ..., t_{N+1}]$. For time step, $\Delta t_n = t_{n+1} - t_n$, we can write a discretized form of equation (4.3):

$$\frac{\Delta\ln(v_j(t_n))}{\Delta t_n} \approx \sum_i^{N_h} \tilde{M}_{ij}h_i(t_n) - m_j, \tag{4.4}$$

where we define the quantitative infection network $\tilde{M}_{ij} \coloneqq M_{ij}\phi_{ij}\beta_{ij}$, and $\frac{\Delta\ln v_j(t_n)}{\Delta t_n} \coloneqq \frac{\ln(v_j(t_{n+1}))-\ln(v_j(t_n))}{t_{n+1}-t_n}$. We can write an analogous equation to equation (4.4) for all time steps and all virus types in the system. All of these equations can be written in a compact form using a single matrix equation:

$$W \approx \left(\tilde{M}^\intercal\, \vec{m}\right) \begin{pmatrix} H \\ \vec{1} \end{pmatrix}, \tag{4.5}$$

where W and H are matrices with elements $W_{ij} = \frac{\Delta\ln v_i(t_j)}{\Delta t_j}$ and $H_{ij} = h_i(t_j)$, $\vec{m}$ is the column vector of decay rates with elements m_i, and $\vec{1}$ is a vector of ones with dimensions $1 \times N$. Given density measurements of the hosts and viruses we can reconstruct the quantitative infection network using equation (4.5). We solve

the following minimization problem to obtain approximations $\tilde{M}_{rec}$ and $\vec{m}_{rec}$ of the quantitative infection matrix, $\tilde{M}$, and the decay rate vector $\vec{m}$:

$$\begin{aligned} \underset{(\tilde{M}^\intercal \vec{m})}{\arg\min} \quad & \left\| W - \left(\tilde{M}^\intercal \vec{m}\right) \begin{pmatrix} H \\ \vec{1} \end{pmatrix} \right\|_2 \\ \text{subject to} \quad & M_{ij} >= 0, \\ & m_i > 0. \end{aligned} \tag{4.6}$$

To solve this problem we used CVX, a package for specifying and solving convex problems [92, 93]. In this study we focus on the reconstruction of the quantitative infection network, but the method also infers decay rates for all virus types in the system. We use a normalized Frobenius distance between the original and reconstructed infection matrices as a metric of the quality of reconstruction, namely:

$$\text{Error}_{\text{rec}} = \frac{\|\tilde{M} - \tilde{M}_{\text{rec}}\|_2}{\|\tilde{M}\|_2}. \tag{4.7}$$

4.3 Results

4.3.1 Reconstruction quality depends on the variability of the dynamics

We begin with an example in which there are 10 host types, 10 virus types and 20 virus-bacteria interactions. The effective infection rates ($\phi * \beta$) vary from 10^{-7} to 5×10^{-6}.Figure 4.1 shows an example of a successful infection network reconstruction using the method described in Section 4.2.3. The matrices W and H were calculated using measurements of the dynamics every 6 min for a total of 96 hours. This results in a reconstruction error $Error_{rec} = 0.01$. The method is able to correctly identify all of the interactions. The small error arises from differences in the inferred quantitative values.

In general, there are multiple factors affecting reconstruction quality. One important factor is the variability of the dynamics. For example, if the dynamics start at

steady state values, there would be no variability in the dynamics, the columns of the matrix H would all be identical and it would not be possible to infer the infection network. We test the effect of variability systematically by performing matrix reconstruction for an ensemble of matrices and different levels of variability. To control variability in the dynamics we change how far the initial densities are from the equilibrium densities. We initialize density of each host and virus type in the system at $x_0 = x_{eq}(1 \pm \delta)$, where x_{eq} is the equilibrium density of a given type and δ is a free parameter that controls the distance from its equilibrium density. We calculated the mean reconstruction error for an ensemble of 100 matrices (Figure 4.2). The reconstruction error has a maximum at $\delta = 0$ (not shown for visualization purposes), which corresponds to starting the system at the equilibrium densities. The quality of the reconstruction increases as the initial conditions move away from the equilibrium densities.

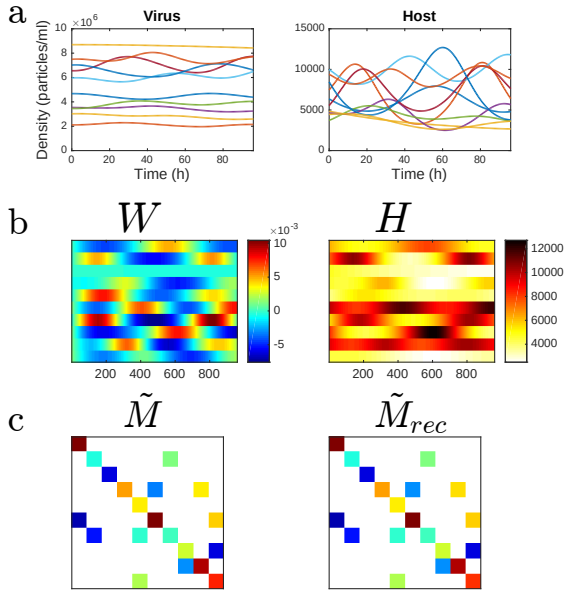

Figure 4.1: Example of infection network reconstruction. (a) Virus and host dynamics for 96 hours. (b) Matrices W and H constructed by taking measurements of virus and host densities every 6 min as described in Section 4.2.3. (c) Original and reconstructed infection matrices ($Error_{rec} = 0.01$). A feasible parameter set was used in the simulation as described in section 4.2.2

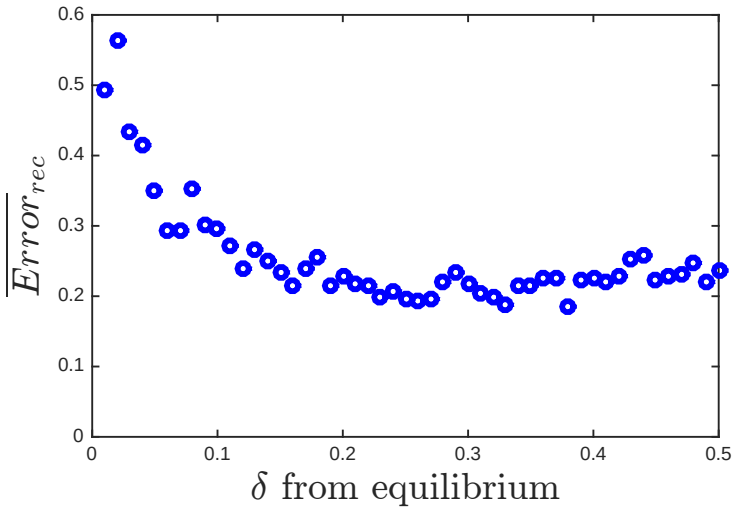

Figure 4.2: Mean reconstruction error as a function of the fraction away from the equilibrium densities, δ, for an ensemble of 100 matrices. Feasible parameter set were used in the simulation as described in section 4.2.2

4.3.2 Reconstruction from multiple experiments: an alternative approach

We propose an improvement to the single experiment approach for reconstruction. In this alternative approach we combine measurements from different experiments to increase reconstruction quality. One key advantage of this approach is that, by increasing the number of experiments used for reconstruction, we can reduce the total time and number of measurements per experiment. This is a crucial advantage in virus-bacteria systems, which are known to evolve rapidly [94, 95, 15]. In the multiple-experiment approach we generate a host matrix H and a virus matrix W by combining matrices from multiple experiments that differ only in their initial conditions (Figure 4.3). This extends equation (4.5) to include information from multiple experiments. Specifically, assuming that we perform p different experiments and calculate matrices $\{H_1, H_2, \ldots, H_p\}$ and $\{W_1, W_2, \ldots, W_p\}$ for each experiment, we can write the system:

$$(W_1 W_2 \dots W_p) \approx \left(\tilde{M}^\intercal \, \vec{m}\right) \begin{pmatrix} H_1 & H_2 & \dots & H_p \\ & & \vec{1} & \end{pmatrix}, \tag{4.8}$$

where $\vec{1}$ is a vector of ones with dimensions $1 \times (N_1 + N_2 + \ldots + N_p)$, assuming that we take N_i measurements from experiment i. Using the same minimization process presented in Section 4.2.3 we can obtain an approximation, $\tilde{M}_{rec}$, of $\tilde{M}$.

Figure 4.4 compares the single and multiple experiments approach for three matrices with different nestedness values. We see how the multiple experiment approach results in lower reconstruction error for the three different cases. Figure 4.5 extends the comparison to an ensemble of 100 different matrices. We compare the multiple experiment approach to the average result of the single experiment approach. For a given matrix we performed 20 different experiments. Each experiment has the same infection matrix and the same model parameters but different initial conditions. We compare the performance of the reconstruction using each experiment individually vs. combining the measurements of the 20 experiments as described in equation (4.8). In this comparison we fix the total number of measurements; We compare the reconstruction error when using 960 measurements from a single experiment (measuring the dynamics every 6 minutes for 96 hours), against the performance when combining the first 48 measurement of all 20 experiments (every 6 minutes for 4.8 hours).

We performed the comparison for 100 different matrices (Figure 4.5). Multiple-experiment reconstruction results in lower error than the average single experiment reconstructions across a wide range of nestedness values. The multiple experiment approach is also more robust; it results in smaller variance in the reconstruction error. Performing more than a few experiments not only decreases the mean reconstruction error, but also decreases the standard deviation significantly (Figure 4.6). For the specific configuration studied here reconstruction error minimizes around 18 experiments.

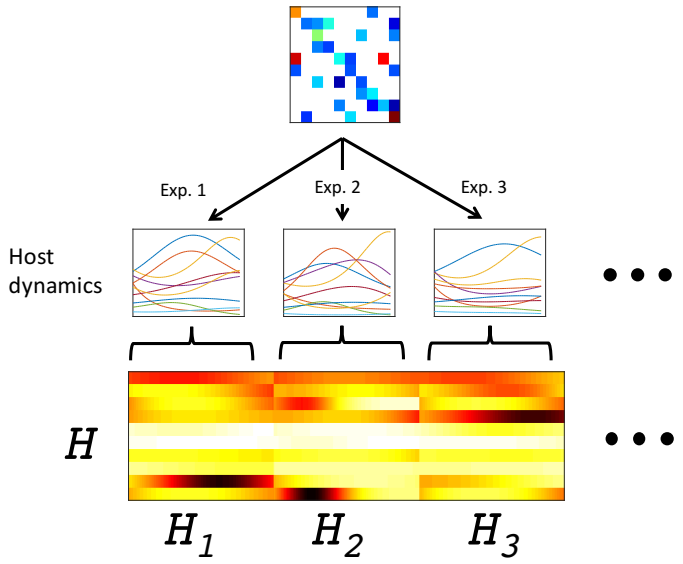

Figure 4.3: Schematic representation of how H is calculated in the multiple-experiment approach. Multiple experiments are performed with the same matrix $\tilde{M}$ and different initial conditions.

4.3.3 Robustness of inference given noise in measurement

Here we evaluate the effect of measurement of white Gaussian noise on the quality of the inference. We follow the same procedure as in the noiseless case to reconstruct infection networks using multiple experiments. Figure 4.7 shows mean reconstruction error for an ensemble of 100 matrices as a function of the signal-to-noise ratio (SNR). We see that using 20 experiments and 48 measurements per experiment, network inference is possible for large signal-to-noise ratio, but reconstruction error increases significantly when the noise approaches 10% of the signal (SNR = 10 dB).

4.4 Discussion

We presented a theory-driven method to estimate host-phage infection networks in an community with multiple virus and host types. Current experimental techniques to measure such networks are difficult to scale to large size systems and most of the

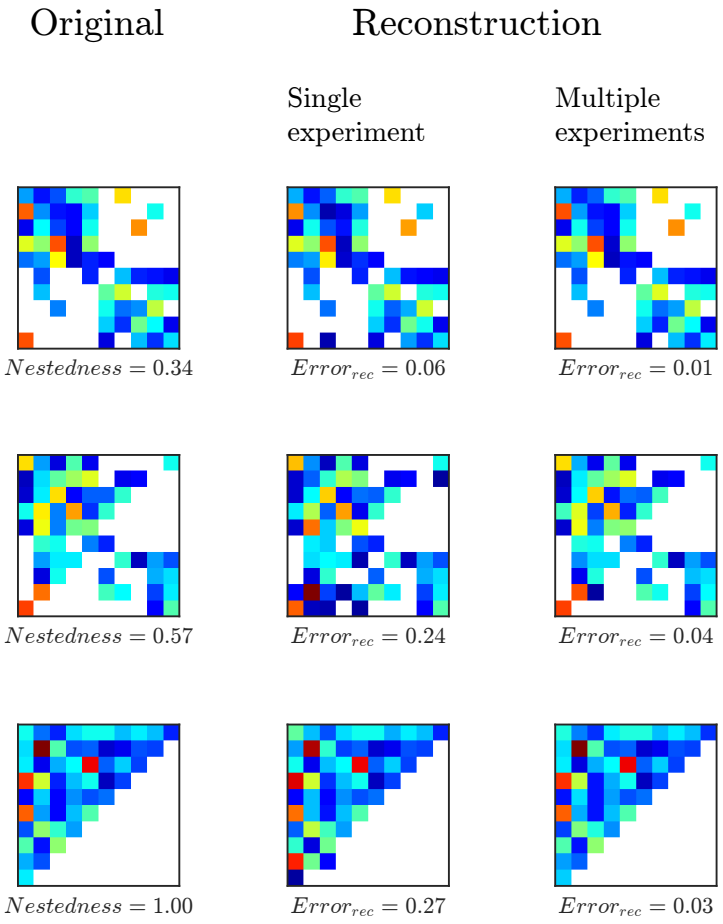

Figure 4.4: Examples of reconstruction for three different matrices and two different methods. Each row shows the original matrix and the resulting reconstruction for each method. The first column shows the original matrices with values of nestedness (NODF): 0.34, 0.55, and 1 respectively. The middle column shows the reconstructed matrices and corresponding reconstruction errors for the single experiment approach using 960 measurements. The last column from the right shows the reconstructed matrices and corresponding errors for the multiple experiment approach using 20 experiments and 48 measurements per experiment. The total number of measurements is the same in the three different methods. The time between measurements is, $\Delta t = 6min$.

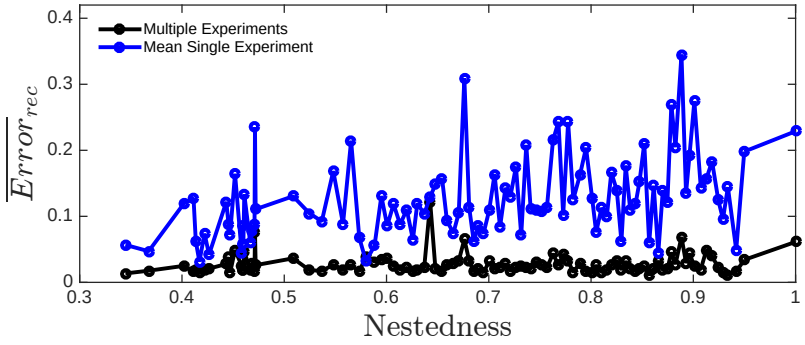

Figure 4.5: Reconstruction error vs Nestedness for two different methods. Black line denotes the reconstruction error, $Error_{rec}$, using the multiple-experiments approach. Blue line describes the mean reconstruction error for the same 20 experiments used in the multiple-experiment approach but using each experiment separately. The total number of measurements is the same in both approaches.

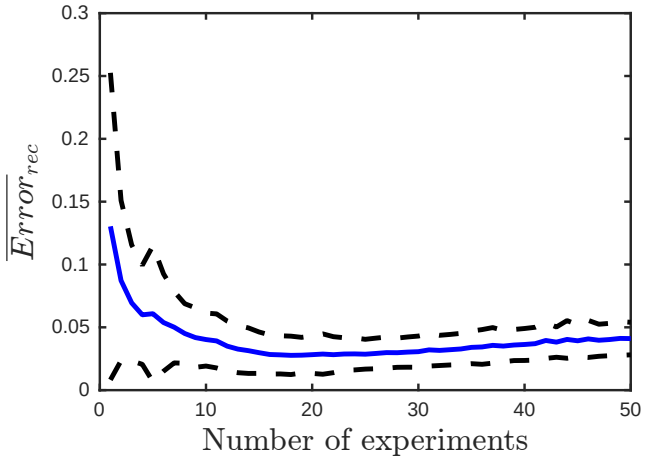

Figure 4.6: Mean (blue line) and standard deviation (dotted line) of the reconstruction error for 100 infection matrices as a function of the number of experiments used in the multiple-experiment approach. Fixed number of total measurements (960). $\Delta t = 6min$.

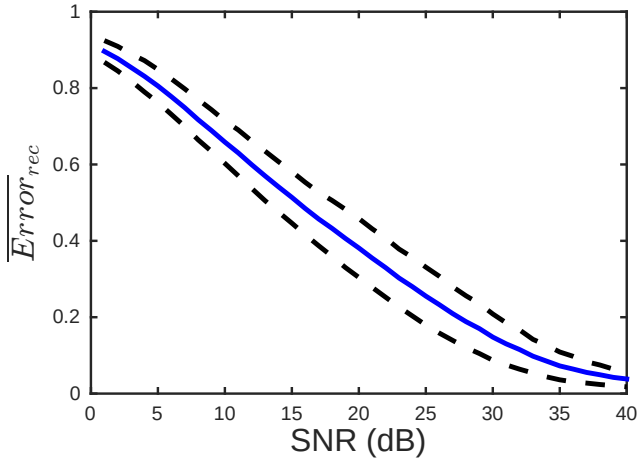

Figure 4.7: Mean (blue line) and standard deviation (dotted line) of the reconstruction error for 100 different matrices as a function of the signal-to-noise ratio. The multiple experiment approach was used to reconstruct the matrix $\tilde{M}$. For each reconstruction, the matrices H and W were constructed using 20 runs with different initial conditions and 48 measurements per run. $\Delta t = 6$ min.

time require virus and/or host isolates, consequently capturing just a subset of all potential interactions present in natural environment. Our approach addresses these limitations by using time-series measurements of experiments involving the whole virus-bacteria community. We also presented an improvement over the single experiment approach for infection network reconstruction. In the multiple-experiment approach we combined measurements from multiple experiments increasing the variability and lowering the reconstruction error. The multiple-experiment approach has the additional advantage of requiring shorter measurement time per experiment. As a consequence, there is a lower probability of a host gaining resistance to a virus type or a virus developing the ability to infect a new host, increasing the chances of reconstructing the infection network of the target community.

The current method takes as input the measured densities of bacteria and phage in an environmental sample. Next-generation high-throughput sequencing techniques

provide a means to characterize bacterial and viral communities in a variety of environmental samples [96, 97, 16, 98, 59]. In the past, such characterization has focused on phylogenetics groups, by using RNA and other marker genes. Such markers are insufficiently resolved with respect to differences in relevant phenotypes, e.g., phage-bacteria infectivity. However, new computational approaches are increasingly available to infer strain, and eventually strain densities, from metagenomic datasets [99, 100]. The increased used of quantitative piplines from sample to strain density for both bacteria and viruses will enable the kind of inference proposed here.

Our present approach uses the equations modeling virus population exclusively, to infer virus-bacteria infection networks. Nonetheless, this method can be expanded by including the equations modeling bacteria population, to infer competitive interactions between bacteria types and bacterial growth rates. The present approach is based on a specific modeling of the interactions in a virus-bacteria communities. Experimental verification is necessary to test whether or not the dynamical model is sufficiently robust representation of naturally occurring systems. Nevertheless, this study presents key steps towards an alternative way of thinking about determining who infects whom in a virus-bacteria community. This view has the potential to significantly reduce the experimental burden, e.g., we are able to infer $n_h \times n_v$ interactions by measuring the dynamics of $n_h + n_v$ organisms, and to overcome the limitations of culture-based approaches by inferring interactions without culturing.

CHAPTER V

CONCLUSIONS

In this thesis we studied the population dynamics of virus-bacteria systems. These complex systems are found everywhere in the planet. It is increasingly obvious that they play a role in shaping the environment we live in. They affect the oceans, the soils, the food we eat, and even our skin and gut. It is easy to see that advances in our understanding of these systems will have important implications to our understanding of the natural world.

The use of mathematical models helped us gain insights into these systems. We were able to obtain testable predictions about the relationship between network structure, life-history traits and biodiversity. We saw how high-dimensional systems with nested infection networks can be maintained through the appropriate trade-offs: the more permissive bacteria have to grow faster than the less permissive ones and viruses with broader host ranges have to be less efficient at exploiting their host than viruses with narrower host ranges. These trade-offs were found for the case of a perfectly nested network and under a specific dynamical model. Naturally occurring systems may differ in the underlying infection network and functional form of the interactions. Nevertheless, our findings point towards a more general trade-off between range of interaction vs effectiveness that can be tested experimentally.

Using the particular case of a perfectly nested network we were able to see the importance of model parameters and their connection to biodiversity. We also saw that this relationship is not as straightforward when studying ensembles of networks. Each particular infection network can me mapped to a specific feasible region of parameter space. Thus, when studying the effect of network structure on biodiversity,

model parameters are entangled in a non-trivial way. Here we showed that averaging over regions of parameter space is not enough to uniquely identify trends of network structure vs biodiversity. We found that studying different regions of parameter space can result in contradictory conclusions about the relationship between nestedness and biodiversity. This conclusions are not specific to the property of nestedness or the host-phage systems studied in this thesis. Instead, they extend to other metrics of network architecture and systems with different interactions. It is not our goal to discourage the use of numerical methods to study the effect of networks architecture on community ecology. The high dimensionality of naturally occurring systems and consequently the number of possible network configurations prohibits exclusively using analytical methods to tackle questions about biodiversity. The ideas presented here are intended to shed light on the importance of studying the parameter space in a systematic way and are complementary to analytical methods for determining biodiversity.

Of course, to study the effect of network structure on biodiversity in an experimental setting we need to have a way of determining infection networks experimentally. Inferring infection networks in a virus-microbe community is not a trivial task, especially in environmental samples. In an effort to overcome some of the limitations of current methods, we presented the theoretical foundation of a novel way of inferring infection networks in environmental samples. The method presented here differs from the standard methods of inferring infection networks in that we exploit the dynamic nature of the interactions in virus-microbe system. We use the effects that organism have on each other over time to infer infectious links. We do this by using discrete measurements of the time-series from experiments where the whole community interact.

In this thesis we tackled some important ecological questions in virus-microbe systems. Nevertheless, we are only addressing a subset of all the complexity in these

systems. Other aspects of virus-microbe communities that we did not directly address point towards possible future directions and potential extension of our results. Here, we studied communities that start with all the possible organisms in them. The composition of the communities were only altered by extinctions of types already in them. In our models we were not accounting for changes in the community as a consequence of evolution. This is a non-trivial to model but important aspect of these communities. Microbial and viral communities are usually engaged in an arms race and are constantly evolving. Considering evolution would allow us to study questions related to the formation and maintenance of infection networks. From this perspective, the insight of this thesis can be considered as operating on snapshots of the evolutionary time scale.

Another important aspect that we can incorporate in our models is space. Our conclusions are based on the assumption of well mixed systems. This assumption can be sensible for some aquatic and laboratory systems. Nevertheless it is well known, that the present of a spacial component can alter ecological outcomes in competitive and predator-prey systems. For example, it is known that by creating niches, space can increase the region of parameter space that allows coexistence in competitive systems. Additionally, virus-microbe systems are seldom found in isolation. They are usually a part of a more complex ecological systems with higher trophic levels that ultimately affect the dynamics of microbes and viruses. The systems presented here also abstract away the resources needed by the host. In natural environment, it is known that these resources interact with virus-microbe systems in complex ways. For example, the lysates resulting from a successful infections can be used again by surviving bacteria in what is known as the"viral shunt".

This theses advances our understanding of virus-microbe complex systems and the relationship between dynamics, life-history traits and network structure. There are still a many known unknowns to be studied and unknown unknowns to be revealed.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 1

A.1 Stability of the boundary fixed points for dynamics with a perfectly nested infection matrix

In this section we show that if conditions (2.4), (2.5), and (2.6) from the main text are satisfied, then all boundary equilibria of the system are unstable. As discussed in the main text, this implies that if the dynamics in the boundary subsystem tend to the boundary equilibrium, then the subsystem can be invaded by at least one of the extinct host or viral strains. If trajectories in the subsystem do not converge to the boundary equilibrium point, then our conclusions about invasability only apply if system (2.1) is permanent (i.e. densities are bounded above and, after some time, are bounded below by a finite value, [48]). When system (2.1) is permanent, invasion at the equilibrium implies invasion along any orbit (over infinite time). This is because the average long term invasability conditions along orbits in permanent Lotka-Volterra systems are equal to the invasability conditions at equilibrium points [48]. We are not aware of any proof that our system is or is not permanent.

For notational convenience, we make two changes to the way the system is presented. First, we don't use the matrix M_{ij} to establish who can infect who. Instead we change the range of sums in accord with the nested pattern. Second, we reverse the order of the hosts ($H_i \to H_{n-i+1}$). In this numbering, host 1 can be infected by all viral strains, and host n can be infected by a single viral strain (Figure A.1). In

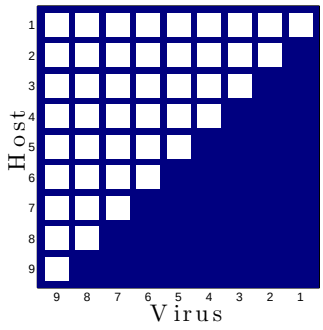

Figure A.1: Perfectly nested infection network with the new host numbering used in A.1.

this new notation the system takes the form:

$$\frac{dH_i}{dt} = r_i H_i \left(1 - \frac{\sum_{j=1}^{n} H_j}{K}\right) - H_i \sum_{j=i}^{n} \phi_j V_j, \tag{A.1}$$

$$\frac{dV_i}{dt} = \phi_i \beta_i V_i \left(\sum_{j=1}^{i} H_j\right) - m_i V_i. \tag{A.2}$$

In this notation, conditions (2.4), (2.5), and (2.6) from the main text become:

$$\frac{m_{i+1}}{\phi_{i+1}\beta_{i+1}} > \frac{m_i}{\phi_i \beta_i} \tag{A.3}$$

$$r_{i+1} < r_i \tag{A.4}$$

$$\frac{m_n}{\phi_n \beta_n} < K. \tag{A.5}$$

Note that, as a consequence of reversing the order of the hosts, condition (A.4) on the growth rates (r_i) is the reverse of condition (2.5).

We will show that all possible boundary fixed points are unstable with respect to invasion. Let $\mathbf{X} = (H_1, H_2, \ldots, H_n, V_1, V_2, \ldots, V_n)$ denote the vector of all host and viral densities and let $\mathbf{X}^* = (H_1^*, H_2^*, \ldots, H_n^*, V_1^*, V_2^*, \ldots, V_n^*)$ denote a fixed point of system (A.1) and (A.2), where H_i^* is the equilibrium density of host i and V_i^* is the equilibrium density of virus i. We will denote the Jacobian by J and the eigenvalues

by λ. We will consider 4 properties of fixed points, such that every possible boundary fixed point has at least one (and possibly more) of these properties. We will show that any fixed points with any of these properties is unstable with respect to invasion by at least one of the host or viral strains. The 4 properties of fixed points are:

1. $\mathbf{X}^*$ where $H_i^* = 0 \quad \forall \quad i \in [1, n]$.

2. $\mathbf{X}^*$ where $H_1^* = H_2^* = H_3^* = \cdots = H_k^* = 0$ with $k < n$ and $H_{k+1}^* \neq 0$.

3. $\mathbf{X}^*$ where $H_i^* = H_{i+1}^* = \cdots = H_n^* = 0$ for $i \neq 1$ and $H_{i-1} \neq 0$.

4. $\mathbf{X}^*$ where $H_i^* = H_{i+1}^* = \cdots = H_k^* = 0$ for $i > 1$, $k < n$, and $H_{k+1} \neq 0$.

The first property corresponds to fixed points where none of the hosts are present. The second property corresponds to fixed points where hosts with low defense level are not present (i.e. hosts that can be infected by all or many viral types are not present). The third one corresponds to fixed points where hosts of high defense are not present (i.e. hosts that can be infected by one or only a few viral types are not present). The last property corresponds to fixed points where host of intermediate defense level are not present.

We will prove the instability of fixed points with properties 1,2,3, and 4 in lemmas 3, 4,5 and 6 respectively. Before that we will prove two more lemmas.

Lemma 1. *Let $\boldsymbol{X}$ be a fixed point, then:*

a) If $H_i^* = 0$, then $\frac{\partial \dot{H}_i}{\partial H_i}\big|_{\boldsymbol{X}^*}$ is an eigenvalue of the Jacobian evaluated at $\boldsymbol{X}^*$.

b) If $V_i^* = 0$, then $\frac{\partial \dot{V}_i}{\partial V_i}\big|_{\boldsymbol{X}^*}$ is an eigenvalue of the Jacobian evaluated at $\boldsymbol{X}^*$.

Proof. a)

Assume $H_i^* = 0$. We can write the equation describing the dynamics of H_i as:

$$\frac{dH_i}{dt} = H_i f(\mathbf{X}), \tag{A.6}$$

where $f(\mathbf{X}^*)$ is a function of all the densities. Then, for any density $X \neq H_i$,

$$\left.\frac{\partial \dot{H}_i}{\partial X}\right|_{\mathbf{X}^*} = \left(H_i \frac{\partial f(\mathbf{X})}{\partial X}\right)\Bigg|_{\mathbf{X}^*} = 0. \tag{A.7}$$

Hence, the Jacobian will have the form:

$$J = \begin{pmatrix} \frac{\partial \dot{H}_1}{\partial H_1} & \cdots & \frac{\partial \dot{H}_1}{\partial H_n} & \frac{\partial \dot{H}_1}{\partial V_1} & \cdots & \frac{\partial \dot{H}_1}{\partial V_n} \\ \vdots & & & & & \vdots \\ 0 \ \ldots \ 0 & \frac{\partial \dot{H}_i}{\partial H_i} & 0 \ \ldots \ 0 & 0 & \ldots & 0 \\ \vdots & & & & & \vdots \\ \frac{\partial \dot{V}_n}{\partial H_1} & \cdots & \frac{\partial \dot{V}_n}{\partial H_n} & \frac{\partial \dot{V}_n}{\partial V_1} & \cdots & \frac{\partial \dot{V}_n}{\partial V_n} \end{pmatrix}_{\mathbf{X}^*} . \tag{A.8}$$

Note that all the elements in the ith row are zero, except for the diagonal term, which has the form $\frac{\partial \dot{H}_i}{\partial H_i}\big|_{\mathbf{X}^*}$. Thus, the characteristic equation can be written as:

$$det(J - I\lambda) = \left(\left.\frac{\partial \dot{H}_i}{\partial H_i}\right|_{\mathbf{X}^*} - \lambda\right) P(\lambda) = 0, \tag{A.9}$$

where $P(\lambda)$ is a polynomial λ. Hence, $\left.\frac{\partial \dot{H}_i}{\partial H_i}\right|_{\mathbf{X}^*}$ is an eigenvalue of the Jacobian.

b) The proof for the case where $V_i^* = 0$ follows the same steps as proof a) for $H_i^* = 0$. □

Lemma 2. *Let $\boldsymbol{X}^*$ be a fixed point where $V_i^* = 0$ for all i and $H_i^* \neq 0$ for at least one i, then $\boldsymbol{X}^*$ is unstable with respect to invasion by at least one virus.*

Proof. Assume $H_i^* \neq 0$. The equilibrium condition for H_i is:

$$r_i \left(1 - \frac{\sum H_j^*}{K}\right) - \cancel{\sum_{j \neq i}^{n} \phi_j V_j^*} = 0. \tag{A.10}$$

So, $\sum H_j^* = K$.

Now, $V_n^* = 0$, so using lemma 1 we know that one eigenvalue of the Jacobian is:

$$\lambda = \left.\frac{\partial V_n^*}{\partial V_n}\right|_{X^*} = \phi_n \beta_n \sum H_j^* - m_n = \phi_n \beta_n K - m_n > 0, \tag{A.11}$$

where the inequality results from the condition of coexistence (A.5). Equation (A.11) implies that the fixed point $\mathbf{X}^*$ is unstable with respect to invasion by virus n. □

Now, we will show that fixed points with any of these 4 properties are unstable.

Lemma 3. *Let $\boldsymbol{X}^*$ be a fixed point where $H_i^* = 0$ for all $i \in [1, n]$, then $\boldsymbol{X}^*$ is unstable with respect to invasion of any of the hosts.*

Proof. Assume $H_i^* = 0$ for all types of hosts in the system, then from equation (A.1) it follows that also $V_i^* = 0$ for all types of viruses at equilibrium.

Using lemma 1,

$$\lambda = \left.\frac{\partial \dot{H}_i}{\partial H_i}\right|_{X^*} = r_i > 0 \tag{A.12}$$

is an eigenvalue. Thus $\mathbf{X}^*$ is unstable with respect to invasion of any of the hosts. □

Lemma 4. *Let $\boldsymbol{X}^*$ be a fixed point where $H_1^* = H_2^* = H_3^* = \cdots = H_k^* = 0$ with $k < n$ and $H_{k+1}^* \neq 0$, then $\boldsymbol{X}^*$ is unstable with respect to invasion of hosts 1 through k .*

Proof. Assume $H_1^* = H_2^* = H_3^* = \cdots = H_k^* = 0$ with $k < n$. The equilibrium condition for V_i for $1 \leq i \leq k$ is:

$$\left(\phi_i \beta_i \sum_{j=1}^{i} H_j - m_i\right) V_i = 0. \tag{A.13}$$

Note that $V_1, V_2, ..., V_k$ only infect hosts 1 through k (the upper limit of the sum is i and $1 \leq i \leq k$). Thus if $H_i^* = 0$ for the first k viruses, then it must be the case that $V_i^* = 0$ for the first k viruses, i.e

$$V_1^* = V_2^* = V_3^* = \cdots = V_k^* = 0. \tag{A.14}$$

From lemma 1 the eigenvalues for the first k hosts have the form :

$$\lambda = \left.\frac{\partial \dot{H}_i}{\partial H_i}\right|_{X^*} = r_i \left(1 - \frac{\sum H_j^*}{K}\right) - \sum_{j=k+1}^{n} \phi_j V_j^*. \tag{A.15}$$

Note the lower limit in the sum as a result of equation (A.14).

Now, assume $H_{k+1} \neq 0$. From the equilibrium condition for H_{k+1} we get:

$$\sum_{j=k+1}^{n} \phi_j V_j^* = r_{k+1} \left(1 - \frac{\sum H_j^*}{K}\right). \tag{A.16}$$

If $V_i^* = 0$ for all i in the sum, then $V_i^* = 0$ for all the viruses in the system. This case is already covered in lemma 2.

On the other hand if $V_i^* \neq 0$ for at least one virus V_i^* we get,

$$\left(1 - \frac{\sum H_j^*}{K}\right) > 0. \tag{A.17}$$

Plugging the result (A.16) in equation (A.15), we get:

$$\lambda = (r_i - r_{k+1}) \left(1 - \frac{\sum H_j^*}{K}\right) > 0, \tag{A.18}$$

where the inequality comes from the inequality (A.17) and the condition for coexistence (A.4). Equation (A.18) implies that the fixed point $\mathbf{X}^*$ is unstable with respect to invasion by host i where $1 \leq i \leq k$.

□

Lemma 5. *Let $\boldsymbol{X}^*$ be a fixed point where $H_i^* = H_{i+1}^* = \cdots = H_n^* = 0$ for $i \neq 1$ and $H_{i-1}^* \neq 0$, then $\boldsymbol{X}^*$ is unstable with respect to invasion of at least one species of host or virus.*

Proof. Assume $H_i^* = H_{i+1}^* = \cdots = H_n^* = 0$. The equilibrium condition for viruses $i-1$ through n are:

$$\begin{aligned}
\left(\sum_{j=1}^{i-1} H_j^* - \frac{m_{i-1}}{\phi_{i-1}\beta_{i-1}}\right) V_{i-1}^* &= 0 \\
\left(\sum_{j=1}^{i-1} H_j^* - \frac{m_i}{\phi_i\beta_i}\right) V_i^* &= 0 \\
&\vdots \\
\left(\sum_{j=1}^{i-1} H_j^* - \frac{m_n}{\phi_n\beta_n}\right) V_n^* &= 0.
\end{aligned} \tag{A.19}$$

Note that all the sums in (A.19) are the same. Because $\frac{m_\ell}{\phi_\ell\beta_\ell} \neq \frac{m_p}{\phi_p\beta_p}$ for all $\ell \neq p$ in the system, the equations in (A.19) can only be satisfied if $V_k^* \neq 0$ for at most one $k \in [i-1, n]$.

To see that $\mathbf{X}^*$ is unstable, consider three possible cases:

Case 1: $\mathbf{X}^*$ where $V_k^* = 0 \quad \forall \quad k \in [i-1, n]$.

Case 2: $\mathbf{X}^*$ where $V_n^* \neq 0$ and $V_k^* = 0 \quad \forall \quad k \in [i-1, n-1]$.

Case 3: $\mathbf{X}^*$ where $V_k^* \neq 0$ for one $k \in [i-1, n-1]$ and $V_\ell = 0$ for $\ell \in [i-1, n]$, $k \neq \ell$.

Case 1:

Assume $V_k^* = 0$ for all $k \in [i-1, n]$. In this case $V_k^* = 0$ for all $k \in [1, n]$. Indeed, the equilibrium condition for H_{i-1} implies:

$$r_{i-1}\left(1 - \frac{\sum H_j^*}{K}\right) - \cancel{\sum_{j=i-1}^{n} \phi_j V_j^*} = 0. \tag{A.20}$$

Thus,

$$\sum H_j^* = K. \tag{A.21}$$

From equation (A.1) we see that equation (A.21) implies $V_i^* = 0$ for all i. This is the case covered in lemma (2).

Case 2:

Assume $V_n^* \neq 0$ and $V_k^* = 0$ for all $k \in [i-1, n-1]$. From lemma 1 we know that, because $V_k = 0$ for $i-1 \leq k < n$ the corresponding eigenvalues have the form:

$$\lambda = \left.\frac{\partial V_k^*}{\partial V_k}\right|_{X^*} = \phi_k \beta_k \sum_{j=1}^{i-1} H_j^* - m_k. \tag{A.22}$$

From the equilibrium condition for V_n, we get:

$$\sum_{j=1}^{i-1} H_j^* = \frac{m_n}{\phi_n \beta_n}. \tag{A.23}$$

Note that the limits of the sums in (A.22) and (A.23) are the same as a result of the H_i^* that are zero, substituting the last expression into (A.22):

$$\lambda = \phi_k \beta_k \frac{m_n}{\phi_n \beta_n} - m_k > 0, \tag{A.24}$$

where the inequality results from (A.3), and consequently the steady state $\mathbf{X}^*$ is unstable with respect to invasion of any of the viruses $i-1$ through $n-1$.

Case 3:

$V_k^* \neq 0$ for one $k \in [i-1, n-1]$ and $V_\ell = 0$ for $\ell \in [i-1, n]$, $k \neq \ell$. The only virus capable of infecting host n is virus n. Using lemma 1 we get that one eigenvalue is given by:

$$\lambda = \left.\frac{\partial \dot{H}_n}{\partial H_n}\right|_{X^*} = r_n \left(1 - \frac{\sum H_j^*}{K}\right), \tag{A.25}$$

where we used $V_n^* = 0$. At this fixed point the only virus infecting host $i-1$ is virus k, so the equilibrium condition becomes:

$$r_{i-1} \left(1 - \frac{\sum H_j^*}{K}\right) - \phi_k V_k^* = 0. \tag{A.26}$$

Thus,

$$\left(1 - \frac{\sum H_j^*}{K}\right) > 0 \tag{A.27}$$

and

$$\lambda = r_n \left(1 - \frac{\sum H_j^*}{K}\right) > 0. \tag{A.28}$$

So, the fixed point $\mathbf{X}^*$ is unstable with respect to invasion by host n. □

Lemma 6. *Let $\boldsymbol{X}^*$ be a fixed point where $H_i^* = H_{i+1}^* = \cdots = H_k^* = 0$ for $i \neq 1$, $k < n$, and $H_{k+1} \neq 0$, then $\boldsymbol{X}^*$ is unstable with respect to invasion of at least one species of host or virus.*

Proof:

Similarly to lemma 5 the equilibrium condition for viruses $i-1$ through k are:

$$\begin{aligned}
\left(\sum_{j=1}^{i-1} H_j^* - \frac{m_{i-1}}{\phi_{i-1}\beta_{i-1}}\right) V_{i-1}^* &= 0 \\
\left(\sum_{j=1}^{i-1} H_j^* - \frac{m_i}{\phi_i\beta_i}\right) V_i^* &= 0 \\
&\vdots \\
\left(\sum_{j=1}^{i-1} H_j^* - \frac{m_k}{\phi_k\beta_k}\right) V_k^* &= 0.
\end{aligned} \tag{A.29}$$

To simultaneously satisfy all of the equations at most one $V_\ell^* \neq 0$ for $\ell \in [i-1, k]$. To show that $\mathbf{X}^*$ is unstable, it is convenient to study two different cases.

Case 1: $\mathbf{X}^*$ where $V_k^* = 0$

Case 2: $\mathbf{X}^*$ where $V_k^* \neq 0$.

Case 1:

Assume $V_k = 0$. Using lemma 1 we know that an eigenvalue of the Jacobian is:

$$\lambda = \frac{\dot{H}_k}{H_k} = r_k \left(1 - \frac{\sum H_j^*}{K}\right) - \sum_{j=k+1}^{n} \phi_j V_j^*. \tag{A.30}$$

The equilibrium condition for H_{k+1} yields:

$$r_{k+1} \left(1 - \frac{\sum H_j^*}{K}\right) - \sum_{j=k+1}^{n} \phi_j V_j^* = 0. \tag{A.31}$$

Using this result the eigenvalue becomes:

$$\lambda = (r_k - r_{k+1}) \left(1 - \frac{\sum H_j^*}{K}\right). \tag{A.32}$$

And as a consequence of condition (A.3), the eigenvalue is positive and the fixed point is unstable with respect to invasion by host k.

Case 2:

Assume $V_k \neq 0$. The equilibrium condition for V_k is:

$$\phi_k \beta_k \sum_{j=1}^{i-1} H_j^* - m_k = 0. \tag{A.33}$$

So,

$$\sum_{j=1}^{i-1} H_j^* = \frac{m_k}{\phi_k \beta_k}. \tag{A.34}$$

From Lemma 1 follows that for any $i \leq \ell < k$, there exists an eigenvalue of the form:

$$\lambda = \left.\frac{\partial V_\ell^*}{\partial V_\ell}\right|_{X^*} = \phi_\ell \beta_\ell \sum_{j=1}^{i-1} H_j^* - m_\ell. \tag{A.35}$$

Substituting equation (A.33), we get:

$$\lambda = \left.\frac{\partial V_\ell^*}{\partial V_\ell}\right|_{X^*} = \phi_\ell \beta_\ell \frac{m_k}{\phi_k \beta_k} - m_\ell > 0, \tag{A.36}$$

where the inequality is a consequence of condition (A.3). Thus, the fixed point is unstable with respect to invasion by viruses $i-1$ through $k-1$.

A.2 Parameters values used in the numerical simulations

Table A.1: Values of the life-history traits used in the numerical simulations. Values based on [1, 2]

		r (h^{-1})	K (cells/ml)	m (h^{-1})	ϕ (ml/(cells×h))	β (No. of viruses)
Fig.2.3	H_1	0.1760	10^7			
	H_2	0.9914	10^7			
	V_1			0.0681	5×10^{-9}	24
	V_2			0.1047	5×10^{-9}	10
Fig.2.4	H_1	1.0500	10^8			
	H_2	0.300	10^8			
	V_1			0.0100	5×10^{-9}	100
	V_2			0.0650	5×10^{-9}	100
Fig.2.5	H_1	0.3000	10^8			
	H_2	1.0500	10^8			
	V_1			0.0100	5×10^{-9}	100
	V_2			0.0650	5×10^{-9}	100
Fig.2.6	H_1	0.3162	10^7			
	H_2	0.6325	10^7			
	H_3	0.8367	10^7			
	H_4	1.1000	10^7			
	H_5	1.1400	10^7			
	V_1			0.0090	5×10^{-9}	20
	V_2			0.0270	5×10^{-9}	40
	V_3			0.0548	5×10^{-9}	60
	V_4			0.0922	5×10^{-9}	80
	V_5			0.1392	5×10^{-9}	100
Fig.2.8b	H_1	0.4472	10^7			
	H_2	0.7583	10^7			
	H_3	0.9747	10^7			
	H_4	1.1511	10^7			
	V_1			0.0100	5×10^{-9}	20
	V_2			0.0375	5×10^{-9}	46
	V_3			0.0650	5×10^{-9}	70
	V_4			0.0925	5×10^{-9}	96

REFERENCES

[1] S. T. Abedon and R. R. Culler, "Bacteriophage evolution given spatial constraint," *J. Theor. Biol.* **248**, 111 (2007).

[2] K. E. Wommack and R. R. Colwell, "Virioplankton: viruses in aquatic ecosystems," *Microbiol. Mol. Biol. Rev.* **64**, 69 (2000).

[3] A. R. Stenholm, I. Dalsgaard, and M. Middelboe, "Isolation and characterization of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*," *Appl. Environ. Microbiol.* **74**, 4070 (2008).

[4] C. O. Flores, J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz, "Statistical structure of host–phage interactions," *P. Natl. Acad. Sci.* **108**, E288 (2011).

[5] Ø. Bergh, K. Y. Børsheim, G. Bratbak, and M. Heldal, "High abundance of viruses found in aquatic environments.," *Nature* **340**, 467 (1989).

[6] C. A. Suttle, "Viruses in the sea," *Nature* **437**, 356 (2005).

[7] C. A. Suttle, "Marine viruses: major players in the global ecosystem," *Nat. Rev. Microbiol.* **5**, 801 (2007).

[8] M. G. Weinbauer, "Ecology of prokaryotic viruses," *FEMS Microbiol. Rev.* **28**, 127 (2004).

[9] J. S. Weitz, *Quantitative Viral Ecology: Dynamics of Viruses and Their Microbial Hosts* (Princeton University Press, 2016).

[10] C. B. Field, M. J. Behrenfeld, J. T. Randerson, and P. Falkowski, "Primary production of the biosphere: integrating terrestrial and oceanic components," *Science* **281**, 237 (1998).

[11] M. R. Clokie, A. D. Millard, A. V. Letarov, and S. Heaphy, "Phages in nature," *Bacteriophage* **1**, 31 (2011).

[12] J. A. Fuhrman, "Marine viruses and their biogeochemical and ecological effects," *Nature* **399**, 541 (1999).

[13] L. F. Jover, T. C. Effler, A. Buchan, S. W. Wilhelm, and J. S. Weitz, "The elemental composition of virus particles: implications for marine biogeochemical cycles," *Nature Reviews Microbiology* **12**, 519 (2014).

[14] S. W. Wilhelm and C. A. Suttle, "Viruses and nutrient cycles in the sea," *BioScience* **49**, 781 (1999).

[15] J. S. Weitz and S. W. Wilhelm, "Ocean viruses and their effects on microbial communities and biogeochemical cycles," *F1000 Biol Rep* **4**, 17 (2012).

[16] R. V. Thurber, M. Haynes, M. Breitbart, L. Wegley, and F. Rohwer, "Laboratory procedures to generate viral metagenomes," *Nature protocols* **4**, 470 (2009).

[17] M. Breitbart, P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer, "Genomic analysis of uncultured marine viral communities," *Proceedings of the National Academy of Sciences* **99**, 14250 (2002).

[18] R. A. Edwards and F. Rohwer, "Viral metagenomics," *Nature Reviews Microbiology* **3**, 504 (2005).

[19] K. Rosario and M. Breitbart, "Exploring the viral world through metagenomics," *Current opinion in virology* **1**, 289 (2011).

[20] A. Campbell, "Conditions for the existence of bacteriophage," *Evolution* , 153 (1961).

[21] B. R. Levin, F. M. Stewart, and L. Chao, "Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage," *American Naturalist* , 3 (1977).

[22] S. T. Abedon, *Bacteriophage ecology: population growth, evolution and impact of bacterial viruses* (Cambridge University Press, Cambridge, UK), p. doi:10.1017/CBO9780511541483.

[23] A. Lotka, "Fluctuations in the abundance of species considered mathematically (with comment by V. Volterra)," *Nature* **119**, 12 (1927).

[24] V. Volterra, "Fluctuations in the abundance of a species considered mathematically," *Nature* **118**, 558 (1926).

[25] T. F. Thingstad, "Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems," *Limnol. Oceanogr.* **45**, 1320 (2000).

[26] C. Winter, T. Bouvier, M. G. Weinbauer, and T. F. Thingstad, "Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited," *Microbiol. Mol. Biol. Rev.* **74**, 42 (2010).

[27] J. S. Weitz *et al.*, "A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes," *The ISME journal* (2015).

[28] U. Bastolla, M. A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, and J. Bascompte, "The architecture of mutualistic networks minimizes competition and increases biodiversity," *Nature* **458**, 1018 (2009).

[29] E. Thébault and C. Fontaine, "Stability of ecological communities and the architecture of mutualistic and trophic networks," *Science* **329**, 853 (2010).

[30] R. M. May, "Will a large complex system be stable?," *Nature* **238**, 413 (1972).

[31] L. F. Jover, M. H. Cortez, and J. S. Weitz, "Mechanisms of multi-strain coexistence in host–phage systems with nested infection networks," *J. Theor. Biol.* **332**, 65 (2013).

[32] C. A. Suttle, "The significance of viruses to mortality in aquatic microbial communities," *Microb. Ecol.* **28**, 237 (1994).

[33] C. Corinaldesi, A. Dell'Anno, M. Magagnini, and R. Danovaro, "Viral decay and viral production rates in continental-shelf and deep-sea sediments of the Mediterranean Sea," *FEMS Microbiol. Ecol.* **72**, 208 (2010).

[34] T. Poisot, G. Lepennetier, E. Martinez, J. Ramsayer, and M. E. Hochberg, "Resource availability affects the structure of a natural bacteria-bacteriophage community," *Biol. Lett.* **7**, 201 (2011).

[35] K. Holmfeldt, M. Middelboe, O. Nybroe, and L. Riemann, "Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts," *Appl. Environ. Microbiol.* **73**, 6730 (2007).

[36] A. Wichels, S. S. Biel, H. R. Gelderblom, T. Brinkhoff, G. Muyzer, and C. Schütt, "Bacteriophage diversity in the North Sea," *Appl. Environ. Microbiol.* **64**, 4128 (1998).

[37] P. Gómez and A. Buckling, "Bacteria-phage antagonistic coevolution in soil," *Science* **332**, 106 (2011).

[38] V. Poullain, S. Gandon, M. A. Brockhurst, A. Buckling, and M. E. Hochberg, "The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage," *Evolution* **62**, 1 (2008).

[39] M. B. Sullivan, J. B. Waterbury, and S. W. Chisholm, "Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*," *Nature* **424**, 1047 (2003).

[40] R. E. Lenski and B. R. Levin, "Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities," *Am. Nat.* , 585 (1985).

[41] J. S. Weitz, H. Hartman, and S. A. Levin, "Coevolutionary arms races between bacteria and bacteriophage," *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9535 (2005).

[42] S. E. Forde, R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst, "Understanding the limits to generalizability of experimental evolutionary models," *Nature* **455**, 220 (2008).

[43] L. M. Childs, N. L. Held, M. J. Young, R. J. Whitaker, and J. S. Weitz, "Multi-scale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin," *Evolution* **66**, 2015 (2012).

[44] A. Weinberger, C. Sun, M. Pluciński, V. Denef, B. Thomas, P. Horvath, R. Barrangou, M. Gilmore, W. Getz, and J. Banfield, "Persisting viral sequences shape microbial CRISPR-based immunity," *PLoS Comput. Biol.* **8**, e1002475 (2012).

[45] J. Haerter and K. Sneppen, "Spatial structure and lamarckian adaptation explain extreme genetic diversity at CRISPR locus," *mBio* **3**, e00126 (2012).

[46] B. R. Levin, "Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria," *PLoS Genetics* **6**, e1001171 (2010).

[47] MATLAB, *version 7.11.584 (R2010b)* (The MathWorks Inc., Natick, Massachusetts, 2010).

[48] J. Hofbauer and K. Sigmund, *Evolutionary games and population dynamics* (Cambridge University Press, Cambridge, UK, 1998).

[49] S. Duffy, P. E. Turner, and C. L. Burch, "Pleiotropic costs of niche expansion in the RNA bacteriophage Φ6," *Genetics* **172**, 751 (2006).

[50] S. Avrani, O. Wurtzel, I. Sharon, R. Sorek, and D. Lindell, "Genomic island variability facilitates Prochlorococcus-virus coexistence," *Nature* **474**, 604 (2011).

[51] C. O. Flores, S. Valverde, and J. S. Weitz, "Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages," *The ISME journal* **7**, 520 (2013).

[52] M. Rosvall, I. B. Dodd, S. Krishna, and K. Sneppen, "Network models of phage-bacteria coevolution," *Phys. Rev. E* **74**, 066105 (2006).

[53] A. Buckling and P. B. Rainey, "Antagonistic coevolution between a bacterium and a bacteriophage," *Proc. Natl. Acad. Sci. U. S. A.* **269**, 931 (2002).

[54] A. Buckling and M. Brockhurst, "Bacteria–Virus Coevolution," *Evolutionary Systems Biology* , 347 (2012).

[55] A. Stern and R. Sorek, "The phage-host arms race: Shaping the evolution of microbes," *Bioessays* **33**, 43 (2011).

[56] B. Kerr, C. Neuhauser, B. J. M. Bohannan, and A. M. Dean, "Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'," *Nature* **442**, 75 (2006).

[57] J. N. Thompson, *The geographic mosaic of coevolution* (University of Chicago Press, 2005).

[58] J. Green and B. J. Bohannan, "Spatial scaling of microbial biodiversity.," *Trends Ecol Evol* **21**, 501 (2006).

[59] F. E. Angly *et al.*, "The marine viromes of four oceanic regions," *PLoS biol* **4**, e368 (2006).

[60] R. M. May, *Stability and complexity in model ecosystems*volume 6 (Princeton University Press, 2001).

[61] A. Roberts, "The stability of a feasible random ecosystem," *Nature* **251**, 607 (1974).

[62] S. Allesina and M. Pascual, "Network structure, predator–prey modules, and stability in large food webs," *Theor. Ecology* **1**, 55 (2008).

[63] J. Bascompte *et al.*, "Disentangling the web of life," *Science* **325**, 416 (2009).

[64] S. Allesina and S. Tang, "Stability criteria for complex ecosystems," *Nature* **483**, 205 (2012).

[65] M. Loreau, *From populations to ecosystems: Theoretical foundations for a new ecological synthesis (MPB-46)* (Princeton University Press, 2010), p. doi:10.1515/9781400834167.

[66] M. Almeida-Neto, P. Guimaraes, P. R. Guimarães, R. D. Loyola, and W. Ulrich, "A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement," *Oikos* **117**, 1227 (2008).

[67] M. E. Newman, "Modularity and community structure in networks," *P. Natl. Acad. Sci. USA* **103**, 8577 (2006).

[68] M. J. Barber, "Modularity and community detection in bipartite networks," *Phys. Rev. E* **76**, 066102 (2007).

[69] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen, "The nested assembly of plant–animal mutualistic networks," *P. Natl. Acad. Sci. USA* **100**, 9383 (2003).

[70] A. James, J. W. Pitchford, and M. J. Plank, "Disentangling nestedness from models of ecological complexity," *Nature* **487**, 227 (2012).

[71] S. Suweis, F. Simini, J. R. Banavar, and A. Maritan, "Emergence of structural and dynamical properties of ecological mutualistic networks," *Nature* **500**, 449 (2013).

[72] D. A. Korytowski and H. L. Smith, "How nested and monogamous infection networks in host-phage communities come to be," *Theor. Ecol.* **8**, 111 (2014).

[73] S. Saavedra, D. B. Stouffer, B. Uzzi, and J. Bascompte, "Strong contributors to network persistence are the most vulnerable to extinction," *Nature* **478**, 233 (2011).

[74] R. P. Rohr, S. Saavedra, and J. Bascompte, "On the structural stability of mutualistic systems," *Science* **345**, 1253497 (2014).

[75] A. D. Tadmor, E. A. Ottesen, J. R. Leadbetter, and R. Phillips, "Probing individual environmental bacteria for viruses by using microfluidic digital PCR," *Science* **333**, 58 (2011).

[76] S. Saavedra and D. B. Stouffer, ""Disentangling nestedness" disentangled," *Nature* **500**, E1 (2013).

[77] A. J. Lotka, *Elements of Physical Biology* (Williams & Wilkins Company, Baltimore, USA, 1925), pp. 92–94.

[78] J. S. Weitz, T. Poisot, J. R. Meyer, C. O. Flores, S. Valverde, M. B. Sullivan, and M. E. Hochberg, "Phage–bacteria infection networks," *Trends Microbiol.* **21**, 82 (2013).

[79] R. Dulbecco and M. Vogt, "Plaque formation and isolation of pure lines with poliomyelitis viruses," *The Journal of experimental medicine* **99**, 167 (1954).

[80] S. Ohno, H. Okano, Y. Tanji, A. Ohashi, K. Watanabe, K. Takai, and H. Imachi, "A method for evaluating the host range of bacteriophages using phages fluorescently labeled with 5-ethynyl-2-deoxyuridine (EdU)," *Applied microbiology and biotechnology* **95**, 777 (2012).

[81] L. Deng, A. Gregory, S. Yilmaz, B. T. Poulos, P. Hugenholtz, and M. B. Sullivan, "Contrasting life strategies of viruses that infect photo-and heterotrophic bacteria, as revealed by viral tagging," *MBio* **3**, e00373 (2012).

[82] E. Allers, C. Moraru, M. B. Duhaime, E. Beneze, N. Solonenko, J. Barrero-Canosa, R. Amann, and M. B. Sullivan, "Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses," *Environmental microbiology* **15**, 2306 (2013).

[83] T. S. Gardner, D. Di Bernardo, D. Lorenz, and J. J. Collins, "Inferring genetic networks and identifying compound mode of action via expression profiling," *Science* **301**, 102 (2003).

[84] M. Hecker, S. Lambeck, S. Toepfer, E. Van Someren, and R. Guthke, "Gene regulatory network inference: data integration in dynamic modelsa review," *Biosystems* **96**, 86 (2009).

[85] A. Arkin and J. Ross, "Statistical construction of chemical reaction mechanisms from measured time-series," *The Journal of Physical Chemistry* **99**, 970 (1995).

[86] T. Hu, A. Leonardo, and D. B. Chklovskii, "Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME)," in *Advances in Neural Information Processing Systems* pp. 790–798 2009.

[87] N. Rubido, A. C. Martí, E. Bianco-Martínez, C. Grebogi, M. S. Baptista, and C. Masoller, "Exact detection of direct links in networks of interacting dynamical units," *New Journal of Physics* **16**, 093010 (2014).

[88] S. G. Shandilya and M. Timme, "Inferring network topology from complex dynamics," *New Journal of Physics* **13**, 013004 (2011).

[89] R. R. Stein, V. Bucci, N. C. Toussaint, C. G. Buffie, G. Rätsch, E. G. Pamer, C. Sander, and J. B. Xavier, "Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota," *PLoS Comput Biol* **9**, e1003388 (2013).

[90] J. Mounier, C. Monnet, T. Vallaeys, R. Arditi, A.-S. Sarthou, A. Hélias, and F. Irlinger, "Microbial interactions within a cheese microbial community," *Applied and environmental microbiology* **74**, 172 (2008).

[91] L. F. Jover, C. O. Flores, M. H. Cortez, and J. S. Weitz, "Multiple regimes of robust patterns between network structure and biodiversity," *Scientific reports* **5** (2015).

[92] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1,", http://cvxr.com/cvx 2014.

[93] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs,", in *Recent Advances in Learning and Control*, edited by V. Blondel, S. Boyd, and H. Kimura Lecture Notes in Control and Information Sciences pp. 95–110 Springer-Verlag Limited 2008.

[94] M. Middelboe, A. Hagström, N. Blackburn, B. Sinn, U. Fischer, N. Borch, J. Pinhassi, K. Simu, and M. Lorenz, "Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria," *Microbial Ecology* **42**, 395 (2001).

[95] J. R. Meyer, D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski, "Repeatability and contingency in the evolution of a key innovation in phage lambda," *Science* **335**, 428 (2012).

[96] G. D. Hannigan and E. A. Grice, "Microbial ecology of the skin in the era of metagenomics and molecular microbiology," *Cold Spring Harbor perspectives in medicine* **3**, a015362 (2013).

[97] E. M. Wood-Charlson, K. D. Weynberg, C. A. Suttle, S. Roux, and M. J. Oppen, "Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise," *Environmental microbiology* **17**, 3440 (2015).

[98] S. Minot, R. Sinha, J. Chen, H. Li, S. A. Keilbaugh, G. D. Wu, J. D. Lewis, and F. D. Bushman, "The human gut virome: inter-individual variation and dynamic response to diet," *Genome research* **21**, 1616 (2011).

[99] J. G. Caporaso *et al.*, "Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms," *The ISME journal* **6**, 1621 (2012).

[100] C. Luo, R. Knight, H. Siljander, M. Knip, R. J. Xavier, and D. Gevers, "ConStrains identifies microbial strains in metagenomic datasets," *Nature biotechnology* **33**, 1045 (2015).